MARKOV CHAINS AT THE INTERFACE OF COMBINATORICS, COMPUTING, AND STATISTICAL PHYSICS

A Dissertation
Presented to
The Academic Faculty

by

Amanda Pascoe Streib

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Algorithms, Combinatorics, and Optimization, School of Mathematics

Georgia Institute of Technology
May 2012

MARKOV CHAINS AT THE INTERFACE OF COMBINATORICS, COMPUTING, AND STATISTICAL PHYSICS

Approved by:

Professor Dana Randall, Advisor
School of Computer Science
Georgia Institute of Technology

Professor Christine Heitsch
School of Mathematics
Georgia Institute of Technology

Professor Eric Vigoda
School of Computer Science
Georgia Institute of Technology

Professor Nina Balcan
School of Computer Science
Georgia Institute of Technology

Professor Prasad Tetali
School of Mathematics and School of Computer Science
Georgia Institute of Technology

Date Approved: 14 March 2012

For my parents,

who always believe in me;

and for Noah,

who never fails to give me confidence and strength

when I have lost my own.

ACKNOWLEDGEMENTS

First of all, I want to thank my advisor, Dana Randall, for all her help and guidance. She always amazes me in her ability to find the key insight to solving hard problems. Her high expectations have helped pushed me and motivated me to be a better researcher and a better writer.

I am grateful for the many opportunities awarded to me at Georgia Tech. I have been fortunate to receive excellent instruction from the world-class faculty in the Algorithms, Combinatorics, and Optimization program. I especially want to thank Prasad Tetali for his support and for introducing me to several stimulating (and sometimes frustrating) problems. Thanks to Luca Dieci for ushering me in and out of this graduate program and for being so approachable and supportive. A special thanks to Klara Grodzinsky, for teaching me how to teach and for answering countless questions with enthusiasm. I also want to thank the ACO program committee for taking a chance on me, and Robin Thomas for encouraging me to join the ACO program. Coming to Georgia Tech was one of the best (and incidentally one of the easiest) decisions I've made. I want to acknowledge the generous support of the National Security Agency through my six year National Physical Sciences Consortium Fellowship.

I also want to thank my professors at Furman for their support and for giving me so many opportunities to grow as a mathematician; in particular, I want to thank Doug Rall for suggesting Georgia Tech for graduate school. My experiences in the Budapest Semesters in Mathematics and in REU's at Central Michigan University, Florida State University, and the Director's Summer Program shaped my trajectory in mathematics and otherwise. I am thankful for the mathematics I learned as well

as the friends and colleagues I met at these programs.

To my fellow graduate students, it has been a pleasure to study with you. I will never forget the long sessions of working through challenging problem sets and studying together for comprehensive exams. I have learned so much from each of you. I have also had the great pleasure to publish joint research with my friends. I especially want to thank Sam Greenberg, Sarah Miracle, and Prateek Bhakta, with whom the work in this thesis was a collaboration.

I want to thank my parents, Frank and Mary Pascoe, for providing me the many opportunities I have had in my career, and my brother, Anthony Pascoe, for being an impressive role model and teacher. Finally, I cannot thank my husband Noah enough for his continual support and unceasing patience with me. I could not have done this without you.

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

The fields of statistical physics, discrete probability, combinatorics, and theoretical computer science have converged around efforts to understand random structures and algorithms. Recent activity in the interface of these fields has enabled tremendous breakthroughs in each domain and has supplied a new set of techniques for researchers approaching related problems. This thesis makes progress on several problems in this interface whose solutions all build on insights from multiple disciplinary perspectives.

First, we consider a dynamic growth process arising in the context of DNA-based self-assembly. The assembly process can be modeled as a simple Markov chain. We prove that the chain is rapidly mixing for large enough bias in regions of Z^d. The proof uses a geometric distance function and a variant of path coupling in order to handle distances that can be exponentially large. We also provide the first results in the case of *fluctuating bias*, where the bias can vary depending on the location of the tile, which arises in the nanotechnology application. Moreover, we use intuition from statistical physics to construct a choice of the biases for which the Markov chain $\mathcal{M}_{mon}$ requires exponential time to converge.

Second, we consider a related problem regarding the convergence rate of biased permutations that arises in the context of self-organizing lists. The Markov chain $\mathcal{M}_{nn}$ in this case is a nearest-neighbor chain that allows adjacent transpositions, and the rate of these exchanges is governed by various input parameters. It was conjectured that the chain is always rapidly mixing when the inversion probabilities are positively biased, i.e., we put nearest neighbor pair $x < y$ in order with bias $1/2 \leq p_{xy} \leq 1$ and out of order with bias $1 - p_{xy}$. The Markov chain $\mathcal{M}_{mon}$ was

known to have connections to a simplified version of this biased card-shuffling. We provide new connections between $\mathcal{M}_{nn}$ and $\mathcal{M}_{mon}$ by using simple combinatorial bijections, and we prove that $\mathcal{M}_{nn}$ is always rapidly mixing for two general classes of positively biased $\{p_{xy}\}$. More significantly, we also prove that the general conjecture is false by exhibiting values for the p_{xy}, with $1/2 \leq p_{xy} \leq 1$ for all $x < y$, but for which the transposition chain will require exponential time to converge.

Finally, we consider a model of colloids, which are binary mixtures of molecules with one type of molecule suspended in another. It is believed that at low density typical configurations will be well-mixed throughout, while at high density they will separate into clusters. This clustering has proved elusive to verify, since all local sampling algorithms are known to be inefficient at high density, and in fact a new nonlocal algorithm was recently shown to require exponential time in some cases. We characterize the high and low density phases for a general family of discrete *interfering binary mixtures* by showing that they exhibit a "clustering property" at high density and not at low density. The clustering property states that there will be a region that has very high area, very small perimeter, and high density of one type of molecule. Special cases of interfering binary mixtures include the Ising model at fixed magnetization and independent sets.

CHAPTER I

INTRODUCTION

The areas of statistical physics, combinatorics, and computer science have converged around efforts to understand random structures and random sampling algorithms. Researchers in each of these fields have been studying similar structures and models, but with different foci, and consequently have developed complementary techniques. For example, statistical physicists, who study random structures associated to physical systems like magnets, use random sampling as a statistical tool to estimate the thermodynamic properties of those systems. In computer science, those same sampling algorithms are used to approximate $\#P$-hard counting problems, such as evaluating the permanent of a matrix. Meanwhile, mathematicians work to classify and count related structures such as matchings in graphs; as we will see shortly, such counting questions are tied closely to sampling those structures. The convergence of these fields has transformed our methods for studying these objects of mutual interest, which now include analytic, experimental, and combinatorial approaches developed in these fields.

A common element of these interdisciplinary studies are Markov chains, which are sampling algorithms that provide a way to estimate statistical properties of random structures. Markov chains are random walks on a set of objects, which eventually converge to a "random" element. They are generally easy to construct and can provide valuable insight, so they are widely used for sampling across the sciences. However, the level of rigor with which the Markov chains are analyzed varies across disciplines. In statistical physics, a Markov chain is considered useful as long as it converges to the correct equilibrium distribution. On the other hand, from a computer scientist's

perspective, Markov chains can only be reliable sampling algorithms if the running time is small enough, and theoretical computer scientists have developed tools for making rigorous guarantees about the efficiency of Markov chains.

We will be especially concerned with issues of efficiency in this thesis. We will describe methods for designing efficient Markov chains based on intuition from physics and show in some cases that algorithms that people believed would work do not. In cases when designing good Markov chains is difficult, we consider alternative approaches, such as analyzing the underlying distributions directly. We have made contributions to all aspects of the study of Markov chains, from solving open problems to developing new tools and utilizing insights from all three of these fields.

In the next few sections, we will give an overview of the role of random sampling in the study of random structures, and we will explore some history about sampling and counting and the use of Markov chains as sampling algorithms. These applications of Markov chains across the sciences are not only good motivation for studying Markov chains, but indeed the applied research in fields such as statistical physics can also provide key insights in the study of their efficiency, as we will see in Section 1.4. Finally, in Section 1.5, we will define the three main problems analyzed in this thesis.

1.1 Random structures and random sampling

Random structures abound in diverse fields of science, including physics, bioinformatics, structural biology, and computer science. The challenge scientists face is to understand these random structures under their (often very complex) probability distributions. For example, the problem of *protein folding* in structural biology is to estimate the three-dimensional structure of a protein given its sequence of amino acids. Biologists do so by taking measurements of the observed probability of each possible fold location and studying the resulting probability distribution over all possible three-dimensional structures [48]. After creating a model that captures the observed

probability distribution, scientists then use tools from probability and statistics to estimate various parameters of their model.

In many cases these questions can be answered by integrating appropriate functions over a set of objects. Consider, for example, the Ising model of ferromagnetism. Here iron atoms are arranged on some graph, say the two dimensional Cartesian lattice $\mathbb{Z}^2$, where each atom is given a positive or negative spin. The energy of a spin configuration is captured by the *Gibbs distribution*, where each configuration σ appears with probability

$$\pi(\sigma) = \frac{1}{Z} e^{-\beta H(\sigma)}, \tag{1}$$

where $H(\sigma)$ is the Hamiltonian, $\beta = 1/T$ is the inverse temperature, and Z is a normalizing constant called the partition function. The Hamiltonian is problem-specific; for the Ising model, the Hamiltonian $H(\sigma)$ is designed to favor configurations where neighboring sites have the same spin. The temperature of the model determines the intensity of the preferences given by the Hamiltonian (a magnet at high temperature loses its magnetism, for example). Information about the thermodynamic properties of the magnets, such as the free energy, specific heat, entropy, mean magnetization per spin, etc., can be obtained by estimating the partition function Z. To calculate Z analytically, physicists would have to evaluate the sum $Z = \sum_{\sigma \in \Omega} e^{-\beta H(\sigma)}$, which is a *discrete integral* of the function $e^{-\beta H(\sigma)}$ over the *state space* Ω, defined as the set of all spin configurations. Analogously, in continuous models the need arises to integrate functions over the state space with respect to the probability measure.

In principal, one can use direct analytic techniques or numerical analysis for evaluating these discrete or continuous integrals. However, for many of these problems, analyzing the distribution directly via counting or integrating can be very challenging and typically all known methods are inefficient. In fact, as we will see in Section 1.2, finding an exact solution for some of these problems is provably intractable. Still, there has been a lot of progress in *estimating* such integrals. In particular, there is

a tremendous body of work in the physics literature toward understanding the Ising model, down to the shape of the clusters of spins that form at very low temperature [21]. Of particular interest in this thesis are efforts to produce relatively simple combinatorial arguments for approximately counting weighted elements in a set, and in fact we provide one such method in our analysis of colloids in Chapter 5.

In practice, however, scientists generally take an experimental approach via random sampling. That is, they design an algorithm which produces samples from the state space according to the desired distribution π. If the algorithm is efficient, then one can run a computer simulation of the algorithm, take random samples from that distribution, and infer properties of a "typical" configuration from the samples obtained. For the Ising model and other physical systems, one can also use random sampling to estimate the partition function, using a sequence of intermediate distributions closely related to the target distribution π; see [42] for details. An important question from a computational perspective is, "how long does it take for the algorithm to converge to the desired distribution?"

While designing a sampling algorithm is often straightforward, providing precise performance guarantees can be quite challenging. For this reason, many sampling algorithms can only be considered heuristics. In fact, if they are used without any rigorous analysis of their efficiency or correctness, then they can give very misleading or incorrect results. Since the late 1980's there has been a push to establish a theoretical underpinning to these heuristics. In this thesis, we prove upper and lower bounds on the running times of such sampling algorithms to better understand their best uses and their limitations.

1.2 Sampling and counting

Once again, calculating the partition function of a physical system is really a question of discrete integration; that is, the goal is to find the sum of the weights over all

possible configurations in the state space Ω. This can be seen as a weighted version of counting. For example, in the monomer-dimer problem from statistical physics, every matching of a graph has a weight associated to it, and estimating the partition function is equivalent to summing the total weight over all matchings of the graph. This generalizes the simple counting problem, "how many matchings does the graph G have?"(which amounts to a sum over all matchings, where every matching of G is given weight 1). Formal counting problems are defined as functions $f : \Sigma \to \mathbb{N}$ from problem instances into the natural numbers. Unfortunately, as alluded to above, for many problems arising in physics and computing, calculating exactly the size of the state space Ω or evaluating an integral of a function over Ω is prohibitively difficult. Often there are no known numerical techniques for evaluating these integrals. In fact, in some cases, it is widely believed that no efficient algorithm exists. To formalize this notion, we need to first establish what we mean by *efficient*.

How does one classify problems in terms of hardness? An insight from computer science is to consider how the running time grows as a function of n, the size of the input. We consider an algorithm efficient if its running time is bounded by a polynomial in n. The field of complexity theory attempts to make rigorous the distinction between problems for which there exists a polynomial-time algorithm (the class P of decision problems) and problems for which we believe that no such algorithm exists (NP-hard problems). See, for example, [38, 59] for more details. Valiant introduced the class #P as an analogue of NP for counting problems [66]. It consists of all counting problems of the form, calculate $f(x)$, for which there exists a polynomial-time nondeterministic Turing machine with exactly f accepting paths. A problem is #P-complete if it is in #P and any function in #P is reducible to f. Generally, the counting version of a problem is more difficult than the decision version. For example, #SAT, which counts the number of satisfying assignments to a formula, is #P-complete. Clearly counting the number of satisfying assignments is at least as

hard as deciding whether there exists a satisfying assignment.

Many problems arising in statistical physics, such as exactly calculating the partition function of the Ising model, are known to be #P-complete [41]. In instances when exact counting is provably intractable (assuming #P≠FP, an analog of P for counting problems), the best we could hope for is an approximation algorithm. A *randomized approximation scheme* for a counting problem f is a randomized algorithm that takes an error parameter $\epsilon > 0$ and for each input $x \in \Omega$, outputs a value N that is within a multiplicative factor of ϵ of the true value $f(x)$ with probability at least 3/4. This algorithm is called a *fully-polynomial randomized approximation scheme*, or *fpras*, if it runs in time that is polynomial in $|x|$ and ϵ^{-1}. In a landmark paper, Jerrum, Valiant, and Vazirani showed that approximate counting and approximate sampling are closely related [43]. In particular, they showed that in order to get an *fpras*, it suffices to construct an *almost uniform sampler*, that is, an approximate sampling algorithm. This underscores the usefulness of Markov chains, which are a very common class of approximate sampling algorithms.

1.3 Markov Chain Monte Carlo algorithms

A Markov chain is simply a random walk on a set of objects, called the *state space*, designed so that after enough iterations, it arrives at a "random" element. For example, shuffling a deck of cards lets you modify the ordering on the cards to get to a random, well-mixed deck (that is, a random permutation). Markov Chain Monte Carlo (MCMC) methods are techniques for sampling using Markov chains with the goal of estimating, for example, partition functions and other multidimensional integrals of functions which are analytically intractable. The basic idea is to estimate an integral (the volume under the curve) by taking random samples from the entire space and then counting the percentage that fall below the curve.

While Markov chains were introduced in classical probability theory, their influence in the scientific community has been much more recent. The first MCMC methods were introduced nearly half a century after Markov developed his probabilistic theory of Markov chains in 1903 [51]. In 1953, Metropolis et al. [56] introduced a Markov chain called the Metropolis algorithm, upon which all other MCMC methods are based. Hastings generalized their results in 1970 [34]. The influence of MCMC methods has since spanned many fields, including statistics, biology, computer science, and optimization. MCMC methods have been used in computer science for approximately counting combinatorial objects (e.g. estimating the permanent [44]) and volume computation [25, 45]. A variation of the Metropolis-Hastings algorithm, called the Gibbs sampler, was introduced in the field of image processing by Geman and Geman in 1984 [31]. With the advent of the modern computer, scientists are finally able to perform reasonable computer simulations of these complex systems using Markov chains, and the use of such simulations is likely to continue to increase.

1.3.1 Markov chains in classical probability theory

In classical probability theory, Markov chains were traditionally studied to understand discrete (or continuous) random processes, with the goal of understanding their long-term behavior. A key distinction between Markov chains and other random processes is that Markov processes are *memoryless*, in the sense that from any given state x, the probability of moving to any state y is independent of all previous positions of the Markov chain. More formally, a Markov chain with transitions $P(\cdot, \cdot)$ over state space Ω is a sequence $X_0, X_1, X_2, \ldots$ of random variables such that for all $x \in \Omega$ and for all $t \geq 0$,

$$P(X_{t+1} = x | X_0, X_1, \ldots, X_t) = P(X_{t+1} = x | X_t).$$

Thus the transition probabilities $P(x, y) = P(X_{t+1} = x | X_t = y)$ can be represented in a stochastic matrix $P = \{P(x, y)\}_{x,y \in \Omega}$, called the *transition matrix*, and the t-step

probability $P^t(x, y)$ is given by the (x, y) entry of the matrix P^t.

The classical analysis of Markov gave sufficient conditions for guaranteeing that a Markov chain would converge to a unique probability distribution in the limit [51]. A distribution π is called a *stationary distribution* of a Markov chain with transition matrix P if $\pi = \pi P$, since once $\pi_0 P^t = \pi$, repeated applications of P do not change the distribution of the Markov chain. A Markov chain is called *ergodic* if it is *irreducible*, i.e. for all $x, y \in \Omega$, there is a t such that $P^t(x, y) > 0$, and *aperiodic*, i.e. for all $x, y \in \Omega$, $\gcd\{t : P^t(x, y) > 0\} = 1$. The following theorem is classical.

Theorem 1.3.1. *Any finite, ergodic Markov chain converges to a unique stationary distribution π; that is, for all $x, y \in \Omega$, $\lim_{t\to\infty} P^t(x, y) = \pi(y)$.*

While this theorem guarantees that an ergodic Markov chain will reach its stationary distribution in the limit, it says nothing about how long that convergence takes.

1.3.2 Computational issues and recent progress

In order for a Markov chain to be useful as a sampling algorithm, it must converge to its stationary distribution quickly. Fortunately, for most applications it is sufficient to wait until the Markov chain is *close to* equilibrium. It is common to measure the distance between the t-step probability distribution and π in terms of the total variation distance

$$\|\mathcal{P}^t(x, \cdot), \pi\|_{tv} = \frac{1}{2} \sum_{y\in\Omega} |\mathcal{P}^t(x, y) - \pi(y)|. \tag{2}$$

Then the efficiency of a Markov chain $\mathcal{M}$ with stationary distribution π can be described by its *mixing time* $\tau(\epsilon)$, defined by

$$\tau(\epsilon) = \max_{x\in\Omega} \min\{t : \|\mathcal{P}^t(x, \cdot), \pi\|_{tv} \leq \epsilon\}, \tag{3}$$

for any $\epsilon > 0$. Then if one is willing to tolerate an error of $\epsilon > 0$, then in order to get a reliable sample, one must run the Markov chain simulation for at least $\tau(\epsilon)$ steps. It is often difficult to decide whether a Markov chain is close to equilibrium.

In practice, scientists often use autocorrelation times and other heuristic estimates of the mixing time [48]. These estimates can be useful but also misleading at times, since it may appear that a Markov chain is close to equilibrium when in fact it is not. On the other hand, if we can provide a theoretical guarantee that the mixing time is at most t, then that algorithm can be used in practice with the assurance that with high probability it will yield a reliable sample as long as it is run for at least t steps. We say that a Markov chain is *rapidly mixing* if $\tau(\epsilon)$ is upper bounded by a polynomial in n and $\log(\varepsilon^{-1})$, where n is the size of each configuration in Ω. On the other hand, if we can prove that the mixing time of an algorithm is at least T, then any experimental results relying on simulations of this algorithm for much fewer than T steps are not likely to be reliable. We call a Markov chain *slowly*, or *torpidly*, *mixing* if the mixing time is at least e^{n^ϵ} for some $\epsilon > 0$.

Aldous and Diaconis were among the first to pioneer methods for rigorously analyzing convergence rates of Markov chains [1, 2]. It was well-known from classical probability theory that the *spectral gap* of a Markov chain is a good measure of its mixing time (see Feller [26], e.g.). The spectral gap is a quantity relating the eigenvalues of the transition matrix; we will define this formally in Chapter 2. Eigenvalues are well-understood and can be computed exactly in many classical applications. Hence, these spectral methods were very useful for analyzing the mixing rates of card shuffling algorithms and random walks on groups [17].

Unfortunately these results are usually not useful for Markov chains we find in applications in physics and computing. Often the transition matrix is exponentially large and it is prohibitively difficult to find the eigenvalues (or even write down the matrix). Fortunately, there has been a lot of progress in developing indirect techniques for bounding the mixing time that are computationally feasible for real-world applications. These include coupling, canonical paths and conductance. For example, in a celebrated paper, Jerrum and Sinclair [39] used conductance to characterize rapid

mixing in terms of the absence of a bottleneck in the state space. We will see details of these techniques and others in Chapter 2. However, there are still many seemingly simple problems for which the known techniques are not sufficient, so there is a great need to develop new methods for bounding the mixing time of Markov chains.

1.4 Markov chains at the interface of physics, combinatorics, and computing

The influence of random sampling algorithms on the world of statistics and related sciences is not a one-way street. In fact, many successes in the study of Markov chains over the last 20 years are due to a new understanding of the connections between computer science, statistical physics, and mathematics. In many cases, researchers in these fields have been studying very similar models. For example, many structures studied in graph theory, such as matchings, colorings, and independent sets, arise naturally in physics and computer science applications. Matchings on a graph correspond to the *monomer-dimer problem* of statistical physics, where the vertices of a lattice are occupied by *monomers* (single-site molecules) and *dimers* (molecules occupying two neighboring sites of the lattice) [42, 35]. Proper q-colorings of a grid graph arise in the zero-temperature Potts model (another model of magnetism), where each site has one of q spins and neighboring sites are forbidden to share the same spin. Independent sets corresponds to the hard-core gas model. This model consists of gas molecules that cannot intersect, where each molecule has a radius larger than the length of an edge between sites, so two neighboring sites cannot both be occupied.

Each of these fields focuses on a unique set of questions and methods for studying the same random structures. However, we are discovering that the answers to these questions can be very enlightening when considered together. For example, there is a surprisingly robust relationship between phase transitions in physical systems and the efficiency of algorithms for sampling those physical systems. Again we will use

the Ising model as an illustrative example. At high temperature a typical configuration contains a fairly random assortment of positive and negative atoms, while at low temperature the influence of neighbors is stronger and the state space will be dominated by configurations which have mostly positive or mostly negative spins. As the temperature is lowered, the Ising model undergoes a phase transition (often called spontaneous magnetization) at the critical temperature T_c. Such phase transitions mark the points at which systems suddenly develop multiple equilibrium (or *Gibbs*) states for infinite-sized lattices [20]. For the Ising model, there is a unique limiting distribution as the size of the lattice tends to infinity at high temperature. However, at low temperature, there are long-range correlations causing the boundary of each state to influence which equilibrium (mostly positive or mostly negative) is reached, even as the boundary goes to infinity. Physicists have developed significant machinery for characterizing the presence of multiple Gibbs states, beginning with Dobrushin in 1968 [20].

A local algorithm (called *Glauber dynamics*) for sampling from the Ising model updates the spin of a single vertex at a time. This algorithm is known to be rapidly mixing at high enough temperature and to require exponential time at low temperature [53, 54, 15, 14]; it is widely believed that the Glauber dynamics transition to slow mixing precisely at the critical temperature. At low temperature, configurations that are mostly positive or mostly negative each have high stationary probability, but to get between such configurations, the Glauber dynamics has to travel through configurations that are evenly divided between positive and negative spins, which have much smaller weight. In this way, the presence of multiple Gibbs states (mostly positive and mostly negative) corresponds to a bottleneck in the state space of the Markov chain which prevents rapid mixing. We will return to this idea in Chapter 2. The situation is similar for several other physical systems, such as independent sets and the Potts model, and their corresponding local dynamics.

Since there is strong evidence that the algorithms' behavior mirrors the behavior of the structures themselves with respect to changes in temperature, there is a clear benefit to taking an interdisciplinary approach when studying these problems. In particular, when studying the Markov chains on these types of systems, it is useful to consider the years of experimental evidence developed in statistical physics. Indeed, many techniques which have proved useful in the study of Markov chains were developed as heuristics in statistical physics. For example, Peierls arguments are combinatorial tools used for proving the existence of multiple Gibbs states, which are also useful to prove that a Markov chain has small conductance [65, 9]. We will use a variation on these methods in our analysis of colloids in Chapter 5. Indeed, there has been a lot of progress made by combining techniques and insight from each of these fields into a coherent, interdisciplinary endeavor.

1.5 Contributions of this thesis

In this thesis, we highlight a few problems in this area for which an interdisciplinary approach was particularly helpful. The examples we consider are fundamental questions of interest across disciplines. First, in Chapter 3, we look at random surfaces arising in self-assembly models in the field of nanoscience. The dynamics of the self-assembly is precisely captured by a local Markov chain, $\mathcal{M}_{mon}$, which samples these surfaces by adding or removing individual squares (or cubes). We consider a biased version of the chain where we are more likely to add a square (or cube) than to remove it. We prove that the chain is rapidly mixing for any uniform bias in Z^2 and for bias $\lambda > d^2$ in Z^d when $d > 2$. Our bounds on the mixing time are optimal on the hypercube. The proof uses a geometric distance function and a variant of path coupling in order to handle distances that can be exponentially large. This is joint work with Sam Greenberg and Dana Randall and appeared in *Proceedings of the Symposium on Discrete Algorithms,* 2009 [33]. In [60], Dana Randall and I extended this analysis

to the more general setting of *fluctuating bias*, where the bias can vary depending on the location of the tile. This work appeared in *Foundations of Nanoscience, 2009*.

Next, in Chapter 4 we will examine the Move Ahead One algorithm for self-organizing lists, which is a fundamental problem in computation that has proved elusive for many years. The Move Ahead One algorithm is a Markov chain $\mathcal{M}_{nn}$ on permutations of n objects whose transitions consist of nearest neighbor transpositions. The chain is known to mix in $\Theta(n^3 \log n)$ time in the unbiased case and $\Theta(n^2)$ time when there is constant bias that favors putting each pair of objects in order rather than out of order. It was conjectured that the chain is always rapidly mixing when the inversion probabilities are positively biased, i.e., we put nearest neighbor pair $x < y$ in order with bias $1/2 \leq p_{xy} \leq 1$ and out of order with bias $1 - p_{xy}$. We prove the chain is always rapidly mixing for two general classes of positively biased $\{p_{xy}\}$. In the first class, p_{xy} depends only on the smaller of x and y. In the second class, the values of p_{xy} are determined by the highest order bit in which x and y differ. Both of these classes include biased permutations with constant bias as a special case. More significantly, we also prove that the general conjecture is false by exhibiting values for the p_{xy}, with $1/2 \leq p_{xy} \leq 1$ for all $x < y$, but for which the transposition chain will require exponential time to converge. This work is joint with Prateek Bhakta, Sarah Miracle, and Dana Randall [6].

Finally, in Chapter 5, we analyze a phenomenon in physics called colloids, in which mixtures of two types of particles tend to separate into clusters of like particles. It is believed that at low density typical configurations will be well-mixed throughout, while at high density they will separate into clusters. This clustering has proved elusive to verify, since all local sampling algorithms are known to be inefficient at high density, and in fact, in previous work we showed that a new nonlocal algorithm requires exponential time in some cases [57]. We define a framework for a special subset of colloids called *interfering binary mixtures*, which contains well-studied models

like the Ising model and Independent Sets. We characterize the high and low density phases for interfering binary mixtures by showing that they exhibit a "clustering property" at high density and not at low density. The clustering property states that there will be a region that has very high area, very small perimeter, and high density of one type of molecule. This is based on joint work with Sarah Miracle and Dana Randall, appearing in *14th International Workshop on Randomization and Computation*, 2011 [58].

CHAPTER II

PRELIMINARIES

In this chapter, we will present some background on Markov chains and techniques for analyzing their mixing times.

2.1 Stationary distributions

Recall from Section 1.3.1 that a Markov chain is ergodic if it is aperiodic and irreducible. Ergodicity is a useful minimum requirement for the Markov chains we construct, since it is sufficient to guarantee that a Markov chain will converge to a unique stationary distribution. However, it is easy to modify a periodic Markov chain to be aperiodic by adding self-loops with probability 1/2 at every state; this removes the periodicity and only slows the chain down by a factor of 2. A Markov chain with self-loops of probability at least 1/2 is called a *lazy* chain.

Given an ergodic Markov chain, the following *detailed balance* condition allows us to easily check if a distribution π is the (unique) stationary distribution of our Markov chain.

Definition 2.1.1. *Let P be the transition matrix for an ergodic Markov chain over a finite state space Ω. If $\pi' : \Omega \to [0, 1]$ is any function satisfying the detailed balance condition:*

$$\pi'(x)P(x, y) = \pi'(y)P(y, x) \tag{4}$$

for all $x, y \in \Omega$, and if it also satisfies $\sum_{x\in\Omega} \pi'(x) = 1$, then π' is the unique stationary distribution of the Markov chain.

As an easy consequence of this condition, we see that if a Markov chain with transition matrix P is to sample from the uniform distribution π, then the Markov chain is

symmetric; that is, $P(x, y) = P(y, x)$ for all $x, y \in \Omega$. This follows from the fact that $\pi(x) = \pi(y)$, since π is the uniform distribution.

It is easy to modify a Markov chain $\mathcal{M}$ which samples from the uniform distribution over the state space Ω to sample from an arbitrary distribution π over Ω. One such method, invented by Metropolis et al. [56] in 1953, uses detailed balance to describe the transition probabilities of the modified chain.

The Metropolis Algorithm

`Starting at any` $x \in \Omega$`, iterate the following:`

- `Choose a transition` (x, y) `according to the Markov chain` $\mathcal{M}$`.`
- `Move to` y `with probability` $\min\left(1, \frac{\pi(y)}{\pi(x)}\right)$.
- `With all remaining probability, stay at` x`.`

Let P be the transition matrix of $\mathcal{M}$ and let $\widehat{P}$ be the transition matrix of the modified chain $\widehat{\mathcal{M}}$. That is, for any pair $x, y \in \Omega$, the probability that $\widehat{\mathcal{M}}$ moves from x to y in one step is $\widehat{P}(x, y) = P(x, y) \min\left(1, \frac{\pi(y)}{\pi(x)}\right)$. Recall that for any pair $x, y \in \Omega$, $P(x, y) = P(y, x)$, since $\mathcal{M}$ samples from the uniform distribution. Therefore, if $\pi(x) \geq \pi(y)$ then

$$\pi(x)\widehat{P}(x, y) \;=\; \pi(x)P(x, y)\frac{\pi(y)}{\pi(x)} \;=\; \pi(y)P(x, y) \;=\; \pi(y)P(y, x) \;=\; \pi(y)\widehat{P}(y, x),$$

and so π satisfies the detailed balance condition and is therefore the unique stationary distribution of the modified chain.

2.2 Mixing times

Thanks to the Metropolis algorithm, it is generally easy to construct an ergodic Markov chain to sample from a given distribution π. However, the more difficult question is how quickly that Markov chain converges to π. In this section, we will review some standard techniques for analyzing the mixing time of a Markov chain.

As discussed in Chapter 1, the spectral gap is a good measure of the mixing time. Let M be a reversible Markov chain with transition matrix P, and label the eigenvalues of P as follows: $1 = \lambda_0 > |\lambda_1| \geq \cdots \geq |\lambda_{|\Omega|-1}|$. Then the *spectral gap* is defined by $\gamma_* = 1 - |\lambda_1|$, and if M is ergodic, $\gamma_* > 0$ (see Section 12.2 of [47], for example). The following theorem relates the spectral gap with the mixing time (Theorems 12.3 and 12.4 [47]):

Theorem 2.2.1. *Let P be the transition matrix of a reversible, ergodic Markov chain with state space Ω, and let $\pi_{min} := \min_{x\in\Omega} \pi(x)$. Then*

$$\frac{1}{2\gamma_*} \log\left(\frac{1}{2\epsilon}\right) \;\leq\; \tau(\epsilon) \;\leq\; \frac{1}{\gamma_*} \log\left(\frac{1}{\epsilon\pi_{min}}\right).$$

Hence to get a bound on the mixing time, it suffices to estimate the spectral gap. When the transition matrix is exponentially large, it can be difficult to calculate its eigenvalues, so we often require more sophisticated methods to bound the mixing time (or the spectral gap).

2.2.1 Conductance

While calculating the spectral gap of a Markov chain directly can be prohibitively difficult, there is a very nice characterization of the spectral gap in terms of the presence or absence of a small cut in the state space. This relationship has been demonstrated using several similar parameters of a Markov chain known as the conductance, the Cheeger constant, and the isoperimetric inequality, each of which provides close upper and lower bounds on the value of the spectral gap. Alon [3] and Alon and Milman [4] related the spectral gap to the "expansion" of a graph. In a similar vein, Lawler and Sokal [46] showed that a discrete version of the Cheeger inequality from differential geometry provides a bound on the spectral gap of a Markov chain. For our purposes, the most convenient version is in terms of the conductance, and is due to Jerrum and Sinclair [40]. The *conductance* of an ergodic Markov chain $\mathcal{M}$ with stationary

distribution π is

$$\Phi_{\mathcal{M}} = \min_{\substack{S \subseteq \Omega \\ \pi(S) \leq 1/2}} \phi_S,$$

where $\phi_S = \phi_S^{(\mathcal{M})}$ is the *conductance of a set* $S \subset \Omega$, defined by

$$\phi_S = \frac{1}{\pi(S)} \sum_{s_1 \in S, s_2 \in \bar{S}} \pi(s_1) P(s_1, s_2).$$

Essentially, S is a bad cut if ϕ_S is exponentially small. The existence of such a set S prevents the Markov chain from mixing rapidly. The important contribution of Jerrum and Sinclair was to show that Φ also provides a lower bound on the spectral gap, so if no such bad cut exists, then the chain is rapidly mixing [40].

Theorem 2.2.2. *For any Markov chain with conductance* Φ *and spectral gap* γ_* *we have*

$$\frac{\Phi^2}{2} \;\leq\; \gamma_* \;\leq\; 2\Phi.$$

2.2.2 Coupling and path coupling

Coupling is widely used in practice as a heuristic to determine whether a Markov chain has converged. A *coupling* of a chain $\mathcal{M}$ is a Markov process on $\Omega \times \Omega$ such that the marginals each agree with $\mathcal{M}$ and, once the two coordinates coalesce, they move in unison thereafter. The idea is that if two copies X_t and Y_t of the Markov chain are run simultaneously, then by the time $X_t = Y_t$, they have lost all influence of the initial configurations X_0 and Y_0 and are therefore close to random. This idea can be made rigorous as follows.

Definition 2.2.1. *Let* $\mathcal{M}$ *be a Markov chain on* Ω*. A* coupling *is a Markov chain on* $\Omega \times \Omega$ *defining a stochastic process* $(X_t, Y_t)_{t=0}^{\infty}$ *such that:*

1. *Each of the processes* X_t *and* Y_t *is a faithful copy of* $\mathcal{M}$ *(that is, e.g.,* $\Pr(X_{t+1} = z | X_t = x, Y_t = y) = P(x, z)$*).*

2. *If* $X_t = Y_t$ *then* $X_{t+1} = Y_{t+1}$*.*

In general, if we run the two Markov chains completely independently, it will take exponential time for them to converge. However, the key is that we do not need the processes to be independent; we can define our coupling in such a way that X_t and Y_t will tend to get closer together, as long as each process, viewed independently, just acts as a copy of the original Markov chain $\mathcal{M}$. The Coupling Lemma bounds the total variation distance probability that the processes have coalesced (see, for example, [1]):

Theorem 2.2.3. $d_{tv}(P^t(x,\cdot),\pi) \leq P(X_t \neq Y_t)$.

Definition 2.2.2. *For initial states x and y, let*

$$T^{x,y} = \min\{t : X_t = Y_t | X_0 = x, Y_0 = y\},$$

and define the coupling time *to be $T = \max_{x,y\in\Omega} E[T^{x,y}]$.*

The following lemma bounds the mixing time in terms of the coupling time of any coupling (see, for example, [1]):

Theorem 2.2.4. $\tau(\epsilon) \leq \lceil Te \ln \epsilon^{-1} \rceil$.

Clearly this bound is only useful if we can define a good coupling, that is, where the expected time to couple is at most a polynomial in n. Path Coupling is a convenient way of establishing this property by only considering a subset of the joint state space $\Omega \times \Omega$. This technique was introduced by Bubley and Dyer [12], but we will use the following version due to Dyer and Greenhill [24].

Theorem 2.2.5. (Dyer and Greenhill [24]) *Let φ be an integer-valued metric defined on $\Omega \times \Omega$ which takes values in $\{0,\ldots,B\}$. Let U be a subset of $\Omega \times \Omega$ such that for all $(x_t, y_t) \in \Omega \times \Omega$ there exists a path $x_t = z_0, z_1, \ldots, z_r = y_t$ between x_t and y_t such that $(z_i, z_{i+1}) \in U$ for $0 \leq i < r$ and $\sum_{i=0}^{r-1} \varphi(z_i, z_{i+1}) = \varphi(x_t, y_t)$. Let $\mathcal{M}$*

be a Markov chain on Ω with transition matrix P and consider any coupling of $\mathcal{M}$. Suppose there exists $\beta \leq 1$ such that

$$\mathbb{E}[\varphi(X_{t+1}, Y_{t+1})] \leq \beta\varphi(X_t, Y_t),$$

for all $t \geq 0$.

1. *If $\beta < 1$, then the mixing time satisfies*

$$\tau(\epsilon) \leq \frac{\ln(B\epsilon^{-1})}{1-\beta}.$$

2. *If $\beta = 1$ (i.e., $\mathbb{E}[\Delta\varphi(X_t, Y_t)] \leq 0$, for all $t \geq 0$, let $\alpha > 0$ satisfy $\Pr[\varphi(X_{t+1}, Y_{t+1}) \neq \varphi(X_t, Y_t)] \geq \alpha$ for all t such that $X_t \neq Y_t$. The mixing time of $\mathcal{M}$ then satisfies*

$$\tau(\epsilon) \leq \left\lceil \frac{\mathrm{e}B^2}{\alpha} \right\rceil \lceil \ln \epsilon^{-1} \rceil.$$

In Chapter 3, we will introduce a distance metric φ that takes on exponentially large values, and in this case we will need a modified version of this Path Coupling Theorem, which we will derive in that chapter.

2.2.3 Comparison of Markov chains

In some cases, it is useful to define a variant of a Markov chain for which we can analyze the mixing time. Then if the new Markov chain is rapidly mixing, we wish to show that the original Markov chain is as well. If P and P' are the transition matrices of two reversible Markov chains on the same state space Ω with the same stationary distribution π, the comparison method (see [18] and [63]) allows us to relate the mixing times of these two chains. Let $E(P) = \{(\sigma, \beta) : P(\sigma, \beta) > 0\}$ and $E(P') = \{(\sigma, \beta) : P'(\sigma, \beta) > 0\}$ denote the sets of edges of the two graphs, viewed as directed graphs. For each σ, β with $P'(\sigma, \beta) > 0$, define a path $\gamma_{\sigma\beta}$ using a sequence of states $\sigma = \sigma_0, \sigma_1, \cdots, \sigma_k = \beta$ with $P(\sigma_i, \sigma_{i+1}) > 0$, and let $|\gamma_{\sigma\beta}|$ denote the length

of the path. Let $\Gamma(\upsilon,\omega) = \{(\sigma,\beta) \in E(P') : (\upsilon,\omega) \in \gamma_{\sigma\beta}\}$ be the set of paths that use the transition (υ,ω) of P. Finally, let $\pi_* = \min_{\rho\in\Omega} \pi(\rho)$ and define

$$A = \max_{(\upsilon,\omega)\in E(P)} \frac{1}{\pi(\upsilon)P(\upsilon,\omega)} \sum_{\Gamma(\upsilon,\omega)} |\gamma_{\sigma\beta}|\pi(\sigma)P'(\sigma,\beta).$$

The following formulation of the comparison method is due to Randall and Tetali [63].

Theorem 2.2.6. *With the above notation, for* $0 < \epsilon < 1$*, we have*

$$\tau(\epsilon) \leq \frac{4\log(1/(\epsilon\pi_*))}{\log(1/2\epsilon)} A\tau'(\epsilon).$$

Other techniques will be developed in the following chapters, where they are needed.

CHAPTER III

SAMPLING BIASED MONOTONIC SURFACES USING EXPONENTIAL METRICS

In this chapter, we examine a Markov chain which simulates the dynamics of a DNA-based self-assembly from nanoscience. The local Markov chain samples surfaces by adding or removing individual squares (or cubes). Since the goal is to create a structure, the chain is biased so that we are more likely to add a square (or cube) than to remove it. We prove that the chain is rapidly mixing for any uniform bias in $\mathbb{Z}^2$ and for bias $\lambda \geq d^2$ in $\mathbb{Z}^d$ when $d > 2$. We also provide the first results in the case of *fluctuating bias*, where the bias can vary depending on the location of the tile. We provide both positive and negative results in this case, showing first how to extend the analysis of the uniform bias case under certain conditions and then showing that when these conditions are violated, the chain may be slowly mixing.

3.1 Sampling monotonic surfaces

Monotonic surfaces spanning finite regions of $\mathbb{Z}^d$ arise in many areas, including nanoscience in the context of DNA-based self-assembly, and combinatorics, in the context of card-shuffling and lozenge tilings. In two dimensions, monotonic surfaces called *staircase walks* are paths on a region of the Cartesian lattice consisting only of steps to the right or down (see Figure 1a). Markov chains for sampling staircase walks can be used to analyze card-shuffling algorithms by associating to a permutation a set of staircase walks [67, 5]. In fact, we will see several bijections of this type in Chapter 4, where we make use of the bounds in this chapter to analyze the mixing time of a natural Markov chain for sampling weighted permutations. *Dyck paths* are

a special case of staircase walks on an $n \times n$ region of $\mathbb{Z}^2$ from $(0, n)$ to $(n, 0)$ which do not cross below the line $x + y = n$; such paths are enumerated by the nth Catalan number C_n. Markov chains for uniformly sampling these paths have been studied, especially in relation to other Catalan structures [55]. In particular, Wilson [67] showed the Markov chain $\mathcal{M}_{mon}$ which at each step selects a peak or valley and inverts it mixes in time $\Theta(n^3 \log n)$ in both the general case and in the case of Dyck paths.

(a) (b)

Figure 1: Monotonic surfaces in two and three dimensions.

Three-dimensional monotonic surfaces correspond to planar tilings or colorings. For example, a lozenge (rhombus) tiling of the triangular lattice can be interpreted visually as a set of unit cubes which are supported on their lower three faces; the boundary of this set of cubes forms a monotonic surface in three dimensions (see Figure 1b). We call this set of cubes σ a *downset*, since the set is downwardly-monotonic, and we let $M(\sigma)$ denote the monotonic surface formed by σ; clearly these are in bijection. Random sampling of lattice configurations like tilings and colorings has attracted attention across many fields. In statistical physics, for example, domino tilings of the Cartesian lattice and lozenge tilings of the triangular lattice are natural models of diatomic molecules and random sampling provides insight into the thermodynamics properties of these systems. Similarly, 3-colorings of lattice regions represent states of the zero temperature Potts model, a popular model of ferro-magnetism. Local Markov chains that update a single site of the coloring, or rotate 2 or 3 tiles in a tiling, can be used to generate random configurations efficiently. For example, the local Markov chain $\mathcal{M}_{mon}$ on planar lozenge tilings that rotates 3 closely packed lozenges by 180 degrees can be interpreted as adding or removing a single cube from

the 3-dimensional structure, as in Figure 1b. Luby, Randall and Sinclair [49] analyzed a family of related, but nonlocal, Markov chains on lozenge tilings and 3-colorings and showed that they are rapidly mixing (or converging rapidly to equilibrium). Subsequently, Randall and Tetali [62] showed that bounds on the mixing times of the nonlocal chains imply that the local chains are also rapidly mixing.

Recently there has been a lot of interest in a *biased* version $\mathcal{M}_{mon}$ of the local Markov chain, where it is more likely to add unit cubes than remove them. More precisely, let $P(\sigma, \tau)$ be the transition probability of moving from σ to τ in one move of the Markov chain $\mathcal{M}_{mon}$. If τ is formed by adding a cube to σ at position $\bar{x}$, then $\lambda_{\bar{x}} = P(\sigma, \tau)/P(\tau, \sigma)$ is the *bias at* $\bar{x}$. If $\lambda_{\bar{x}} > 1$ for every $\bar{x}$, the stationary distribution favors configurations with more cubes. This biased version of the Markov chain arises in nanoscience [50] and biased card shuffling [5], while a biased version of the chain for sampling 3-colorings comes up in the context of asynchronous cellular automata [16]. We restrict our attention here to the first two examples where the monotonic surfaces correspond to sets of supported cubes, although the methodology also applies in the third setting.

Previous work has focused on the case where the biases are *uniform*; that is, $\lambda_{\bar{x}} = \lambda$ for every $\bar{x}$, for which the stationary probability will be proportional to $\lambda^{|\sigma|}$, where $|\sigma|$ is the number of unit cubes in σ. In two dimensions, the uniform bias Markov chain is equivalent to an asymmetric simple exclusion process, which Benjamini et al. [5] studied in order to analyze a biased card shuffling algorithm that favors putting each pair of cards in the lexicographically correct order. They give a bound of $O(n)$ on the mixing rate of the biased chain on $h \times w$ regions of $\mathbb{Z}^2$ (where $hw = n$) for any uniform bias $\lambda > 1$ that is constant with respect to n. These bounds are optimal when $h = w = \sqrt{n}$. In three and higher dimensions, substantially less is known. Majumder et al. [50] show that the chain mixes quickly when the uniform bias is $\Omega(n)$; apparently the case of large bias is the most interesting for nanotechnology

applications. Nothing else is known about the convergence of the biased chain; both of these results are highly technical and do not readily generalize to other values of the bias or other dimensions.

3.1.1 Our results

We make progress in several aspects of the problem of sampling biased surfaces. First, we consider the case of uniform bias. In two dimensions we show that the biased chain is rapidly mixing for any uniform bias on a large family of simply-connected regions, even when the bias is arbitrarily close to one. Our proof is significantly simpler than the arguments of Benjamini et al., while achieving the same optimal bounds on the mixing time for square regions when the bias is constant. In fact, on rectangular $h \times w$ regions of $\mathbb{Z}^2$, where $h \leq w$, we get improved bounds of $O(w(h + \ln w))$, which is optimal when $h = \Omega(\ln w)$. Specifically, we prove the following theorem for rectangular regions:

Theorem 3.1.1. *Let R be a rectangular $h \times w$ region in $\mathbb{Z}^2$ with uniform bias $\lambda \geq 1$. Suppose without loss of generality that $h \leq w$. Let $\chi = \lambda - 1$. Then*

1. *If $\chi > 0$, then the mixing time of $\mathcal{M}_{mon}$ satisfies $\tau(\varepsilon) = O\left(\chi^{-2} w(h + \ln w) \ln \varepsilon^{-1}\right)$.*
2. *If $\chi \geq 0$, then $\tau(\varepsilon) = O\left(w^3 (h + \ln w)^2 \ln \varepsilon^{-1}\right)$.*

If $\lambda > 1$ is a constant and $h = \Theta(w)$, then part 1 of this theorem applies and gives a bound of $O(n)$, where $n = hw$ is the area of the region. On the other hand, for λ very close to 1, part 2 provides a polynomial bound on the mixing time. We prove similar bounds for all *nice* regions, to be defined in Section 3.4.1.2; essentially, these regions are simply-connected and have no holes.

In higher dimensions, we show the chain is rapidly mixing on d-dimensional lattice regions provided the bias $\lambda \geq d^2$. Again, our bounds on the mixing time are optimal when the regions are hyper-cubes, and show the chain is rapidly mixing for a large family of simply-connected regions. For hypercubes we show:

Theorem 3.1.2. *Let R be the d-dimensional $h \times h \times \cdots \times h$ hypercube with volume $h^d = n$ and uniform bias $\lambda \geq d^2$. Then the mixing time of $\mathcal{M}_{mon}$ satisfies $\tau(\varepsilon) = O\left(n \ln \varepsilon^{-1}\right)$.*

The key observation underlying these results is that there is an exponential metric on the state space such that the distance between pairs of configurations is always decreasing in expectation. We then show how to modify the path coupling theorem to handle the case when the distances are exponentially large and the expected change in distance is small during moves of the coupled chain. We believe that this new theorem is of independent interest.

Moreover, we introduce the first results for the more general setting of *fluctuating bias*, where the bias at $\bar{x}$ depends on $\bar{x}$. This situation is more realistic in the context of self-assembly, since several properties of the growth processes can affect the rates of attaching and detaching, and it has been noted that previous methods that allow us to analyze systems with uniform bias do not readily generalize to the fluctuating setting. First, we show that the Markov chain $\mathcal{M}_{mon}$ is rapidly mixing provided the biases do not differ by too much using a similar exponential distance metric. We also use a simple hitting time argument to show rapid mixing in a $d-$dimensional hypercube whenever the minimum bias λ_L satisfies $\lambda_L \geq d$:

Theorem 3.1.3. *Let R be the d-dimensional $h \times h \times \cdots \times h$ hypercube with volume $h^d = n$ and fluctuating bias. Assume the minimum bias λ_L satisfies $\lambda_L \geq d$. Then the mixing time of $\mathcal{M}_{mon}$ satisfies $\tau(\varepsilon) = O\left(h^{2d-1} \ln \varepsilon^{-1}\right) = o(n^2 \ln \varepsilon^{-1})$.*

Notice that this yields a polynomial bound on the mixing time for a uniform bias of $\lambda \geq d$.

From our results for the uniform bias, we can infer rapid mixing for the fluctuating bias chain whenever the minimum bias λ_L satisfies the conditions of the uniform bias mixing theorem (Theorem 3.4.2) and the maximal tiling has constant probability in

the stationary distribution. In particular, in two dimensions we get the following theorem:

Theorem 3.1.4. *Let R be a rectangular $h \times w$ region in $\mathbb{Z}^2$ with fluctuating bias. Suppose the minimum bias λ_L is a constant larger than 1. Suppose without loss of generality that $h \leq w$. Then the mixing time of $\mathcal{M}_{mon}$ satisfies $\tau(\varepsilon) = O\left(w(h + \ln w) \ln \varepsilon^{-1}\right)$.*

Again, this yields the optimal mixing time of $O(n)$ if R is a square.

Given these results, it is tempting to suspect that $\mathcal{M}_{mon}$ is rapidly mixing for any choice of the $\{\lambda_x\}$ as long as each $\lambda_x > 1$. However, we illustrate an example that contradicts this conjecture to prove the following theorem:

Theorem 3.1.5. *There exists a set of biases $\{\lambda_x\}$ on an square region of $\mathbb{Z}^2$ where $\lambda_x > 1$ for all x and yet the mixing time of $\mathcal{M}_{mon}$ is at least $e^{n^{1/3}-1}/4 - 1/2$.*

The remainder of the chapter is organized as follows. In Section 3.2 we formalize the model and the Markov chain. In Section 3.3 we review the path coupling method and introduce the modified path coupling theorem that is more appropriate when distances are exponentially large. In Section 3.4, we show how we can use our new path coupling theorem and introduce the hitting time argument to conclude that the chain on biased surfaces is rapidly mixing. In Section 3.5, we generalize these techniques to apply to the setting of fluctuating bias. Finally, in the conclusion, we discuss other related problems, such as 3-colorings.

3.2 The model

The nanoscience application that motivates this work is a model of DNA-based self-assembly (see, e.g., [30, 64, 68, 69]). Roughly "square" shaped tiles are constructed from strands of DNA so that each side of the tiles is single-stranded. Certain pairs of tiles are encouraged to line up and attach along edges by encoding corresponding sides with complementary sequences of base pairs. At appropriately chosen temperatures, these tiles will have a good chance of assembling according to these prescribed

rules, although they also have a chance of disassociating and breaking apart. The model considered by Majumder et al. [50] allows the left column and bottom row of a large square to first form, and then iteratively allows tiles to associate with the large substrate if their left and bottom neighbors are already present (see Figure 1a). Likewise, tiles can disassociate if their upper and right neighbors are not present, although disassociation happens at a lower rate. The dynamics of this model are precisely captured by the local Markov chain $\mathcal{M}_{mon}$ on 2-dimensional monotonic surfaces and the chain must be rapidly mixing if the substrate is to efficiently self-assemble, as required. The 3-dimensional analogue is also used to study self-assembly, where now tiles are shaped like cubes (as in Figure 1b) and complementary sequences are used to encourage corresponding faces to attach.

In this chapter we study the dynamics of this self-assembly in terms of the monotonic surfaces that form the boundary of the growing substrate. In this section, we formalize these monotonic surfaces and describe the Markov chain in detail.

3.2.1 Monotonic surfaces

While our analysis holds for a general family of simply-connected regions in $\mathbb{Z}^d$ that we call *nice*, we first consider monotonic surfaces forming over simple, rectangular regions R in $\mathbb{Z}^2$. The generalization is straightforward, but requires some careful notation, so for simplicity we postpone the details until Section 3.4.1.2. In $\mathbb{Z}^2$, a *monotonic surface (or path)* in R is a path starting and ending on the boundary of R that only takes steps down or to the right and is composed entirely of edges with both endpoints in R. Such a path is illustrated in Figure 1a when R is a 4×4 square. Notice that any monotonic surface can be interpreted as the upper boundary of a set of unit squares (which we call a downset), where each square in the set is supported below or to the left by other squares in the set or the boundary of R.

Recall that M is a bijection between downsets σ and their corresponding monotonic surfaces $M(\sigma)$. We define the state space Ω_{mon} to be the set of all downsets of R.

3.2.2 The biased Markov chain

We are now ready to describe the Markov chain on Ω_{mon}. For simplicity, we start by defining the *unbiased* chain $\widehat{\mathcal{M}}_{mon}$ that converges to the *uniform distribution* over monotonic paths Ω_{mon}. Start at an arbitrary downset, e.g., let $\sigma_0 = R_L$, where R_L is the empty downset, and repeat the following steps. If we are at a downset σ_t at time t, pick a diagonal d that is parallel to the vector $\overline{u}^* = (1, 1)$ and which intersects the monotonic path at a vertex v. Also, pick an integer $b \in \pm 1$ uniformly at random. If $b = +1$, add the cube above and to the right of the vertex v to create σ_{t+1}, if this is a valid downset. If $b = -1$, let σ_{t+1} be obtained from σ_t by removing the cube below and to the left of v if this is a valid downset. In all other cases, keep σ_t unchanged so that $\sigma_{t+1} = \sigma_t$.

Lemma 3.2.1. *For any rectangular region R, the Markov chain $\widehat{\mathcal{M}}_{mon}$ connects the state space Ω_{mon}.*

Proof. Let σ be any downset and let x^{max} be any cube in σ such that $\sum_i x_i^{max}$ is maximized, if it exists. We can always remove x^{max} and move to $\sigma' = \sigma \setminus x^{max}$ without violating the downset condition. Thus, from any valid downset σ we can always remove points and get to the "lowest" downset R_L. Also, such a sequence of steps can be reversed to move from R_L to any other downset ρ. □

Since we have shown that the moves of $\widehat{\mathcal{M}}_{mon}$ connect the state space and all valid moves have the same transition probabilities, we can conclude from detailed balance (Definition 2.1.1) that the chain converges to the uniform distribution over downsets Ω_{mon}. We now define the *biased* Markov chain by using Metropolis-Hastings probabilities (refer to Section 2.1 or see [56] for details) to modify the transition

probabilities so that we converge to the desired distribution on biased surfaces. This new chain connects the state space by the same argument as in Lemma 3.4.1.

The Biased Markov chain $\mathcal{M}_{mon}$

Starting at any σ_0, iterate the following:

- Choose a diagonal d and a bit $b \in \{\pm 1\}$ as described above.
- If $b = +1$, add the cube above and to the right of the vertex v to create σ_{t+1}, if this is a valid downset.
- If $b = -1$, the with prob. $\frac{1}{\lambda_x}$ let σ_{t+1} be obtained from σ_t by removing the cube x below and to the left of v if this is a valid downset.
- Otherwise let $\sigma_{t+1} = \sigma_t$.

The biased Markov chain $\mathcal{M}_{mon}$ converges to the correct distribution on Ω_{mon} by the detailed balance condition. Moreover, notice that for any given d there is at least one choice of b that proposes a move (adding or removing) that does not result in a valid downset. Therefore $\mathbb{P}[\sigma_{t+1} = \sigma_t] \geq 1/2$, so $\mathcal{M}_{mon}$ is a lazy chain.

3.3 Path coupling with exponential metrics

Recall from Section 2.2.2 that path coupling is a standard technique used to bound mixing times. Although a naive application of it will not work here, we will see that with some new ideas, we can make it work. One of the innovations behind our proofs is to introduce a new metric, and in some cases this metric requires a modified Path Coupling theorem. We present our new theorem here.

To understand why it is difficult to use coupling to prove that $\mathcal{M}_{mon}$ is rapidly mixing, we first examine the straightforward coupling of (σ_t, ρ_t) in the uniform bias case. The natural coupling simply chooses the same diagonal d and bit b to generate

both σ_{t+1} and ρ_{t+1}. We first consider a natural distance metric on $\Omega_{mon} \times \Omega_{mon}$ called the Hamming distance, where $h(\sigma_t, \rho_t) = |\sigma_t \oplus \rho_t|$ (and $\oplus$ is the symmetric difference). However, with this coupling and metric, we face difficulty with even the simplest of pairs (σ_t, ρ_t).

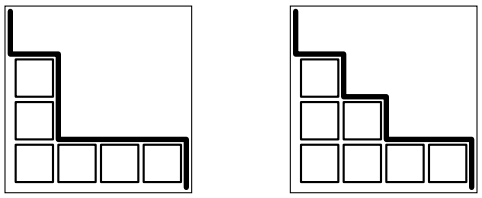


Figure 2: A pair of downsets σ_t (left) and ρ_t (right) where $\rho_t = \sigma_t \cup \{(2,2)\}$.

Examine the pair of downsets in Figure 2. They differ on a single point, so $h(\sigma_t, \rho_t) = 1$. In order to use the coupling theorem above, the expected distance $\mathbb{E}[h(\sigma_{t+1}, \rho_{t+1})]$ must be at most $h(\sigma_t, \rho_t)$. For this pair of downsets, there are two moves that decrease that distance; if $\mathcal{M}_{mon}$ chooses the diagonal $d_0 = \{(0,0) + t\overline{u}^* : t \geq 0\}$ and either $b = +1$ or $b = -1$, then $(\sigma_{t+1}, \rho_{t+1})$ is (ρ_t, ρ_t) or (σ_t, σ_t), respectively. In either case, the distance between σ_{t+1} and ρ_{t+1} decreases by 1. There are also two moves that *increase* the distance. If $\mathcal{M}_{mon}$ chooses $d = \{(1,0) + t\overline{u}^* : t \geq 0\}$ and $b = +1$ or $d = \{(0,1) + t\overline{u}^* : t \geq 0\}$ and $b = +1$, then ρ_{t+1} gains a new point ($(3,2)$ or $(2,3)$, respectively), but σ_{t+1} remains unchanged; no addition to σ_t of a vector along that diagonal leaves a valid downset. With either of these choices, the distance between σ_{t+1} and ρ_{t+1} increases by 1. If $\lambda = 1$, this is sufficient for coupling; the expected change in distance is 0. Unfortunately, for any $\lambda > 1$, the two bad moves happen with probability 1, whereas the two good moves happen with probability 1 and $1/\lambda$, respectively. Therefore the expected distance between the pair (σ_t, ρ_t) *increases* after one step. In higher dimensions, the situation becomes even worse. For the pair of 3 dimensional downsets in Figure 6, there are *three* moves which increase the Hamming distance and only two moves which decrease the distance. Of course, the three moves that increase the distance succeed with probability 1, but one of the

two moves which decreases the distance only succeeds with probability $1/\lambda$.

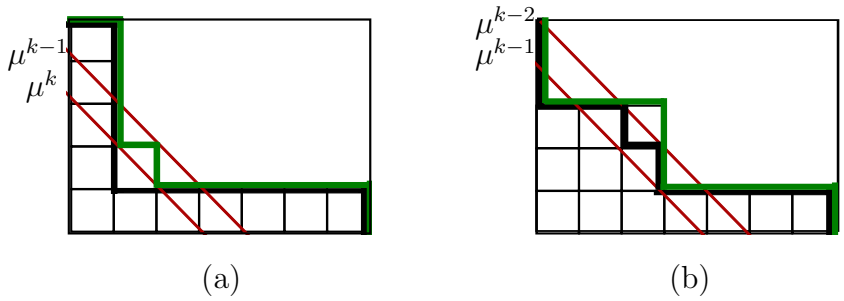


Figure 3: Path coupling with an exponential distance metric.

One promising remedy is to alter the distance metric. The bad cases described above involve two downsets that differ on some point x, where the two moves which decrease the distance involve removing x from $\sigma_t \oplus \rho_t$, while the moves that increase the distance involve adding $x + \overline{u}_i$ to $\sigma_t \oplus \rho_t$ for some i (where $\overline{u}_1 = (1, 0)$ and $\overline{u}_2 = (0, 1)$). Since the bad moves happen with greater probability than the good moves, we consider a distance metric that counts the distance between two sets that differ on x as greater than the distance between two sets that differ on $x + \overline{u}_i$. Specifically, we give a different weight to each Northwest-Southeast diagonal, with the weights smaller along higher diagonals (as in Figure 3a). This allows us to make the change in distance nonpositive in the above cases. Of course we must ensure that the difference in weight is not too great. This is because the opposite situation might happen as well, where the two bad moves involve removing $x - \overline{u}_i$ for some i, (as in Figure 3b); although this situation was not a problem for the Hamming distance metric, if we assign too much weight to those bad moves, the change in distance might be positive in this case. We find the following distance metric suffices. First, let R be the $h \times w$ rectangle in $\mathbb{Z}^2$ and define $\mu = \sqrt{\lambda} \geq 1$. Then for two downsets σ, ρ in R, let

$$\phi(\sigma, \rho) = \mu^{w+h} \sum_{x \in \sigma \oplus \rho} \mu^{-\|x\|_1}, \tag{5}$$

where $\| \cdot \|_1$ is the L_1 norm. Notice that all elements on each Northwest-Southeast diagonal have the same L_1 norm, and so this metric assigns a weight of μ^k for the kth

diagonal from the top right, as in Figure 4. Notice also that this definition ensures that the distance between any two downsets is either 0 or at least 1. We present the proof that this metric is decreasing in expectation at every step in Section 3.4.1.

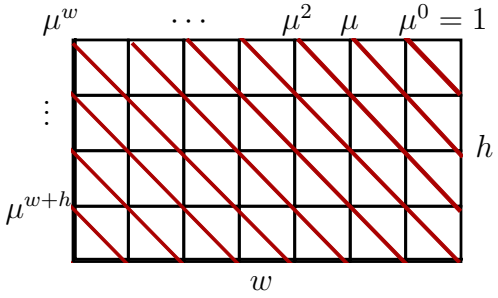


Figure 4: Exponential distance metric.

Unfortunately, this definition of the distance metric presents new problems. First, the distances might now take on non-integer values, while the Path Coupling Theorem requires integer valued metrics. In fact, if this restriction is merely removed, then the theorem is no longer true as the distances might get smaller and smaller without coalescence occurring in a polynomial number of steps. However, it is enough to add the additional condition that no pairs of configurations can have a distance within the open interval $(0, 1)$. The second, more serious concern is that the maximum distance between two configurations can be exponentially large in n. If the distance only changes by a small (polynomial) amount in each step, then we cannot expect the distance to be zero in only a polynomial number of steps. For example, for small λ we can find configurations x_t and y_t so that $\mathbb{E}[\varphi(x_{t+1}, y_{t+1})] \leq (1 - 2^{-n})\varphi(x_t, y_t)$, so the expected change is too small to apply the first part of Theorem 2.2.5. Moreover, the maximum distance B is very large, so we cannot get a good bound on the mixing time using the second part of Theorem 2.2.5 either.

The following modification of the Path Coupling Theorem allows us to handle cases when the distances can be exponentially large and the expected change in distance is small (or even zero). We show that it suffices to prove that the expected change in the absolute value of the distance is proportional to the current distance, and with

this condition the mixing time is polynomially bounded. We apply this new theorem to the biased Markov chain $\mathcal{M}_{mon}$ in Section 3.4.1.

Theorem 3.3.1. *Let ϕ be a metric defined on $\Omega \times \Omega$ which takes finitely many values in $\{0\} \cup [1, B]$. Let U be a subset of $\Omega \times \Omega$ such that for all $(X_t, Y_t) \in \Omega \times \Omega$ there exists a path $X_t = Z_0, Z_1, \ldots, Z_r = Y_t$ such that $(Z_i, Z_{i+1}) \in U$ for $0 \leq i < r$ and $\sum_{i=0}^{r-1} \phi(Z_i, Z_{i+1}) = \phi(X_t, Y_t)$.*

Let $\mathcal{M}$ be a lazy Markov chain on Ω and let (X_t, Y_t) be a coupling of $\mathcal{M}$, with $\phi_t = \phi(X_t, Y_t)$. Suppose there exists $\beta \leq 1$ such that, for all $(X_t, Y_t) \in U$,

$$\mathbb{E}[\phi_{t+1}] \leq \beta \phi_t.$$

1. *If $\beta < 1$, then the mixing time satisfies*

$$\tau(\varepsilon) \leq \frac{\ln(B\varepsilon^{-1})}{1-\beta}.$$

2. *If there exists $\kappa, \eta \in (0,1)$ such that $\mathbb{P}\left[|\phi_{t+1} - \phi_t| \geq \eta \phi_t\right] \geq \kappa$ for all t provided that $X_t \neq Y_t$, then*

$$\tau(\varepsilon) \leq \left\lceil \frac{3e \ln^2(B)}{\ln^2(1+\eta)\kappa} \right\rceil \left\lceil \ln \varepsilon^{-1} \right\rceil.$$

There are two important differences between Theorem 2.2.5 and Theorem 3.3.1. The first is that Theorem 3.3.1 allows for non-integer metrics (provided that for all $X, Y \in \Omega$, $\phi(X,Y) < 1$ implies $\phi(X,Y) = 0$). This is a minor restructuring of the proof of Theorem 2.2.5 [49], and follows exactly from their method. The second is that β may equal 1 while B is exponentially large; this is the case in which both parts of Theorem 2.2.5 were unable to prove rapid mixing. This can be shown again with a slight modification of the original proof, essentially replacing the original distance $\phi(X_t, Y_t)$ with $\ln(\phi(X_t, Y_t))$. There are some technical details concerning the expectation and variance of the logarithm, but the novelty of Theorem 3.3.1 is more in the statement of the result than a new method of proof.

Note that including this case of $\beta = 1$ and exponential B requires a strong bound on the variance of ϕ_t. Without this bound on variance, Theorem 3.3.1 is not true; if $\phi_0 = 2^n$ and $\phi_{t+1} = \phi_t - 1$ for all $t \geq 1$, then clearly it will take time exponential in n for $\phi_t = 0$.

In order to prove Theorem 3.3.1 to handle exponential metrics, we define a new variable ψ, which is essentially $\ln(\phi)$. However, if we hope to prove rapid mixing by looking at $\ln(\phi)$, we need to bound the time to reach $\ln(0) = -\infty$, and the expected time could be unbounded. In particular, in order to prove rapid mixing, we need that the sequence $\{\psi_t\}$ has bounded differences. The technical fix that we make relies on the assumption that $\phi_t \notin (0, 1)$, so we need only bound the time until we reach a negative value for $\ln(\phi_t)$. Hence we define

$$\psi_t = \begin{cases} \ln(\phi_t) & \text{if } \phi_t > 0 \\ -2\ln 2 & \text{if } \phi_t = 0 \end{cases}.$$

This means that $\psi_t \in [-2\ln 2, \ln B]$. The particular value at zero is chosen so that if the expected distance ϕ_t is non-decreasing, then the expected value of ψ_t is non-decreasing, and that if the variance of ϕ_t is at least a linear factor, then the variance of ψ_t is at least a constant.

The following Martingale Lemma follows the proof of Lemma 6 in [49].

Lemma 3.3.2. *Given any bounded function $\phi(t)$, with $d \leq \phi(t) \leq D$ for some $d, D \in \mathbb{R}$ and for all $t \geq 0$, and a stopping value q, let $T = \min\{t : \phi(t) = q\}$. If, for all $t \geq 0$, we have $\mathbb{E}[\phi(t+1) - \phi(t)] \leq 0$ and $\mathbb{E}[(\phi(t+1) - \phi(t))^2] \geq Q$ for some Q, then*

$$\mathbb{E}[T] \leq \frac{2D^2 + q^2 - 2qD}{Q}.$$

Proof. Define the process $Z(t) := (D - \phi(t))^2 - Qt$. Examining the expected difference

between $Z(t)$ and $Z(t+1)$, we have

$$\begin{aligned}\mathbb{E}[Z(t+1) - Z(t)] &= \mathbb{E}[(D - \phi(t+1))^2 - (D - \phi(t))^2] - Q \\ &= \mathbb{E}[-2D(\phi(t+1) - \phi(t)) + \phi(t+1)^2 - \phi(t)^2] - Q \\ &= 2(\phi(t) - D)\mathbb{E}[\phi(t+1) - \phi(t)] + \mathbb{E}[(\phi(t+1) - \phi(t))^2] - Q \geq 0.\end{aligned}$$

Also, since the differences $Z(t+1) - Z(t)$ are bounded, so $\{Z(t)\}$ is a submartingale. T is a stopping time for $Z(t)$, so we may apply the Optional Stopping Theorem for submartingales to deduce that

$$\begin{aligned}\mathbb{E}[T] &\leq \frac{1}{Q}\left[\phi(0)(2D - \phi(0)) + q^2 - 2qD\right] \\ &\leq \frac{2D^2 + q^2 - 2qD}{Q}.\end{aligned}$$

□

Now we may prove the exponential metric theorem, Theorem 3.3.1.

Proof. **Part (1)** This case follows directly from the proof of Theorem 2.2.5, while allowing non-integer valued metrics. Since $\mathbb{E}[\phi_{t+1}] \leq \beta\phi_t$ for all t, it follows that

$$\mathbb{E}[\phi(X_t, Y_t)] = \mathbb{E}[\phi_t] \leq \beta^t\phi_0 = \beta^t\phi(X_0, Y_0) \leq \beta^t B. \tag{6}$$

Since ϕ_t is nonnegative, takes values in $\{0\} \cup [1, B]$ and is equal to zero whenever $X_t = Y_t$, we have

$$\begin{aligned}\mathbb{E}[\phi_t] &= \int_1^B x\Pr(\phi_t = x)dx \\ &\geq \int_1^B \Pr(\phi_t = x)dx \\ &= \Pr(\phi_t \geq 1) \\ &= \Pr(X_t \neq Y_t).\end{aligned}$$

Then since by Equation 6, $\mathbb{E}[\phi_t] \leq \epsilon$ whenever $t \geq \log(B\epsilon^{-1})/\log(\beta^{-1})$, the Coupling Lemma 2.2.3 implies

$$\tau(\epsilon) \leq \ln(B\epsilon^{-1})/\ln(\beta^{-1}).$$

Since $\ln(\beta^{-1}) > 1 - \beta$, Part (1) follows.

Part (2) For part (2), we will show that ψ_t satisfies the conditions of Lemma 3.3.2, with $q = -2\ln 2$, $D = \ln B$ and $Q = \ln(1+\eta)^2\kappa$. Note that this proves the theorem, since Theorem 2.2.4 implies that

$$\tau(\epsilon) \leq \lceil \mathbb{E}[T] e \ln \epsilon^{-1} \rceil \leq \left\lceil \frac{2D^2 + q^2 - 2qD}{Q} e \ln \epsilon^{-1} \right\rceil = \left\lceil \frac{3e(\ln B)^2}{\ln(1+\eta)^2\kappa} \right\rceil \lceil \ln(\epsilon^{-1}) \rceil.$$

First we show that since $\mathbb{E}[\phi_{t+1} - \phi_t] \leq 0$, then also $\mathbb{E}[\psi_{t+1} - \psi_t] \leq 0$. We may assume $\phi_t \neq 0$. Given the value of ϕ_t, let $\{r_0, r_1, r_2, \ldots, r_N\}$ be the possible values for ϕ_{t+1}, each occurring with probability $\{\zeta_0, \zeta_1, \zeta_2, \ldots, \zeta_N\}$. That is, $\mathbb{P}[\phi_{t+1} = r_i | \phi_t] = \zeta_i$, with $\sum_{i=0}^{N} \zeta_i = 1$. Assume $r_0 = 0$. As our chain is lazy, $\mathbb{P}[\phi_{t+1} = \phi_t] \geq 1/2$. Therefore $\zeta_0 \leq 1/2$. Now,

$$\begin{aligned}
\mathbb{E}[\psi_{t+1} | \psi_t] &= \zeta_0(-2\ln 2) + \sum_{i=1}^{N} \zeta_i \ln_2(r_i) \\
&= -2\ln 2\zeta_0 + \ln\left(\prod_{i=1}^{N} r_i^{\zeta_i}\right) \\
&\leq -2\ln 2\zeta_0 + \ln\left(\frac{\sum_{i=1}^{N} \zeta_i r_i}{1-\zeta_0}\right) \\
&= \ln\left(\mathbb{E}[\phi_{t+1} | \phi_t]\right) - 2\ln 2\zeta_0 - \ln(1-\zeta_0) \\
&\leq \ln\left(\mathbb{E}[\phi_{t+1} | \phi_t]\right) \\
&\leq \ln \phi_t \;=\; \psi_t,
\end{aligned}$$

where the first inequality is by the Arithmetic-Geometric Mean Inequality, and the second follows from the fact that $\ln(1-\zeta_0)/\zeta_0 \geq -2\ln 2$ for $\zeta_0 \in (0, \frac{1}{2})$.

Next we prove that if there exist constants $\kappa, \eta \in (0,1)$ such that $\mathbb{P}[|\phi_{t+1} - \phi_t| \geq \eta\phi_t] \geq \kappa$ for $\phi_t \neq 0$, then

$$\mathbb{E}[|\psi_{t+1} - \psi_t|] \geq \ln(1+\eta)\kappa + \ln 2\mathbb{P}[\phi_{t+1} = 0].$$

Let $\zeta_0 = \mathbb{P}[\phi_{t+1} = 0]$. Then

$$\begin{aligned}\kappa &\leq \mathbb{P}[|\phi_{t+1} - \phi_t| \geq \eta\phi_t] \\ &= 1 \cdot \zeta_0 + \mathbb{P}[|\phi_{t+1} - \phi_t| \geq \eta\phi_t | \phi_{t+1} \neq 0](1 - \zeta_0).\end{aligned}$$

Now because, by definition, $\psi_{t+1} = \ln(\phi_{t+1})$ when $\phi_{t+1} \neq 0$, we have

$$\begin{aligned}&\mathbb{P}[|\phi_{t+1} - \phi_t| \geq \eta\phi_t | \phi_{t+1} \neq 0] \\ &\quad= \mathbb{P}\left[\frac{\phi_{t+1}}{\phi_t} - 1 \geq \eta | \phi_{t+1} \neq 0\right] + \mathbb{P}\left[\frac{\phi_{t+1}}{\phi_t} - 1 \leq -\eta | \phi_{t+1} \neq 0\right] \\ &\quad= \mathbb{P}\left[\psi_{t+1} - \psi_t \geq \ln(1+\eta) | \phi_{t+1} \neq 0\right] + \mathbb{P}\left[\psi_{t+1} - \psi_t \leq \ln(1-\eta) | \phi_{t+1} \neq 0\right] \\ &\quad\leq \mathbb{P}\left[|\psi_{t+1} - \psi_t| \geq \ln(1+\eta) | \phi_{t+1} \neq 0\right].\end{aligned}$$

Since $\phi_t \geq 1$, we have $\psi_t \geq 0$, so $|-2\ln 2 - \psi_t| \geq 2\ln 2$. Note that $m := \ln^2(1+\eta) < \ln^2(2)$, since $\eta < 1$. This yields

$$\begin{aligned}\mathbb{E}[(\psi_{t+1} - \psi_t)^2] &= (-2\ln 2 - \psi_t)^2 \zeta_0 + \sum_{\ell \in \Omega, \ell \neq 0} \left(\ln(\ell) - \ln\phi_t\right)^2 \mathbb{P}[\phi_{t+1} = \ell] \\ &\geq (2\ln 2)^2 \zeta_0 + m\mathbb{P}\left[|\psi_{t+1} - \psi_t| \geq m | \phi_{t+1} \neq 0\right](1 - \zeta_0) \\ &\geq (2\ln 2)^2 \zeta_0 + m\left(\frac{\kappa - \zeta_0}{1 - \zeta_0}\right)(1 - \zeta_0) \\ &= \left((2\ln 2)^2 - m\right)\zeta_0 + m\kappa \\ &> m\kappa + 3\zeta_0 \ln^2 2.\end{aligned}$$

Hence we have $Q = \ln(1+\eta)^2\kappa \leq \mathbb{E}[(\psi_{t+1} - \psi_t)^2]$, as desired, so we may apply Lemma 3.3.2 to get

$$E[T] \leq \frac{2\ln^2 B}{\ln^2(1+\eta)\kappa}.$$

□

3.4 Fast mixing of the uniform bias Markov chain

In this section, we examine the uniform bias Markov chain $\mathcal{M}_{mon}$. In Section 3.4.1, we use our exponential metric to show that $\mathcal{M}_{mon}$ is rapidly mixing whenever $\lambda \geq d^2$ for

arbitrary dimension d, and for all $\lambda > 1$ when $d = 2$. In Section 3.4.2, we will present a simple hitting time argument that proves the biased chain converges in polynomial time as long as the minimum bias is at least d and the region is a $d-$dimensional hypercube. We conjecture that the chain is rapidly mixing for all values of a uniform bias $\lambda > 1$ in all dimensions d, but do not yet have a proof for small values of λ in dimensions higher than 2.

3.4.1 Exponential metric for the uniform bias chain

First, we will handle the simple case when R is a rectangular region in $\mathbb{Z}^2$, and then in Section 3.4.1.2, we will explain how to extend this analysis to more complex regions and higher dimensions.

3.4.1.1 Rectangular regions in two dimensions

Next we see how to use the distance metric given in Equation 4 to bound the mixing time of $\mathcal{M}_{mon}$, proving Theorem 3.1.1.

We return to the coupling of (σ_t, ρ_t) that simply supplies the same diagonal d and bit b to both σ_t and ρ_t. We let U be the set of pairs of downsets that differ on a single cube. However, instead of the Hamming distance, we use the distance metric given in equation 4:

$$\phi(\sigma, \rho) = \mu^{w+h} \sum_{x \in \sigma \oplus \rho} \mu^{-\|x\|_1},$$

where $\mu = \sqrt{\lambda} \geq 1$. We will show that this distance metric satisfies non-negative contraction in ϕ_t, which is one of the requirements for Theorem 3.3.1. However, before we can prove that the distances decrease on average, we examine the moves which can *increase* the distance.

For a pair $(\sigma_t, \rho_t) \in U$, there are two different ways the distance can increase in $(\sigma_{t+1}, \rho_{t+1})$. If $\sigma_t = \rho_t \cup \{x\}$, we can increase the distance by attempting to add a cube v that succeeds in σ_t but fails in ρ_t, as in Figure 5a. This occurs when $v = x + \overline{u}_i$

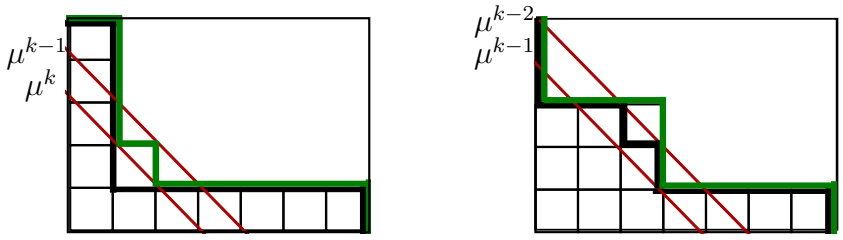


Figure 5: Path coupling with an exponential distance metric. Two cases where $\sigma_t = \rho_t \cup \{x\}$.

for some i, so v is "supported" in σ_t but not ρ_t. Notice that the distance metric ϕ gives these bad moves weight that is μ times smaller than the weight of the two good moves (adding or removing x), counteracting their higher probability. The other way to increase the distance between σ_t and ρ_t is to remove a v that succeeds in ρ_t but not in σ_t. This occurs when $v = x - \overline{u}_i$ for some i, as the move creates a valid downset in ρ_t but not in σ_t, as in Figure 5b. In this case, the distance metric ϕ gives these bad moves weight that is μ times larger than the weight of the two good moves, but for small enough μ, they will still not outweigh the good moves, since the bad moves are less likely to occur than the good moves in this case. We may now prove Theorem 3.1.1.

Proof. We will show that the distance metric ϕ defined above satisfies the conditions of Path Coupling Theorem for Exponential metrics, Theorem 3.3.1. First we want to show that the expected change in ϕ is negative. There are at most 2 choices of (d, b) that can increase ϕ_t. We claim that each of these has an expected increase of at most $\phi_t \lambda^{-1/2}$. To see this, consider a move of the form $v = x + \overline{u}_i$ for some i. Then the increase in distance is $(\sqrt{\lambda})^{w+h-\|v\|_1} = \phi_t \lambda^{-1/2}$. If the move is of the form $v = x - \overline{u}_i$ for some i, then the increase in distance is $\lambda^{-\|v\|_1/2} = \phi_t \sqrt{\lambda}$, but the chance of choosing an appropriate p is $1/\lambda$. Therefore the expected increase is at most $\phi_t \lambda^{-1/2}$.

There are also two choices of (d, b) that *decrease* ϕ_t; corresponding to adding x

and removing x. These each decrease ϕ_t by ϕ_t, and succeed with probability 1 and $1/\lambda$, respectively. Let $\alpha = h + w$; this is the number of choices of diagonals d. The expected change in distance satisfies

$$\begin{aligned}\mathbb{E}[\phi_{t+1} - \phi_t] &\leq \frac{1}{2\alpha}\left(d \cdot \frac{\phi_t}{\sqrt{\lambda}} - \left(1 + \frac{1}{\lambda}\right)\phi_t\right) \\ &= \frac{\phi_t}{2\alpha}\left(\frac{d}{\sqrt{\lambda}} - 1 - \frac{1}{\lambda}\right) \\ &\leq -\frac{\phi_t \chi}{2\alpha}.\end{aligned}$$

Next we check the other conditions of Theorem 3.3.1. For arbitrary $\sigma, \rho \in \Omega_{mon}$, if $x \in \sigma \oplus \rho$ for some x, then $\phi(\sigma, \rho) \geq \sqrt{\lambda}^{h+w-\|x\|_1} \geq 1$. Therefore if $\phi(\sigma, \rho) < 1$, $\phi(\sigma, \rho) = 0$. Let U be the set of pairs of downsets that differ on a single vector. For arbitrary $\sigma, \rho \in \Omega_{mon}$, we can connect σ to ρ by simply adding or removing the vectors in $\sigma \oplus \rho$ one by one, and $\phi(\sigma, \rho)$ is the sum of the distances. Since the volume of R is n, there are at most n possible cubes in $\sigma \oplus \rho$, so $\phi(\sigma, \rho) \leq n\lambda^{h/2}$ for all σ, ρ.

We consider two cases. If $\lambda - 1 > \chi$, then $\mathbb{E}[\phi_{t+1}] \leq \beta\phi_t$, where $\beta = \Theta(1/\alpha)$. Thus, by Theorem 3.3.1, we have $\tau(\epsilon) = O(\chi^{-2} w(h + \ln w) \ln \epsilon^{-1})$. On the other hand, if $\lambda - 1 > 0$, but less than any constant, then we use the second part of Theorem 3.3.1. For any pair of σ, ρ, $\mathcal{M}_{mon}$ can always add a vector v^* in their difference that maximizes $\|v\|_1$. This would change ϕ_t by at least ϕ_t/α. The appropriate v^* is chosen with probability at least $1/\alpha$ and the appropriate b is chosen with probability $1/2$ (and every p succeeds when adding). Therefore there is a $1/(2\alpha)$ chance of changing ϕ_t by $\phi_t/(2\alpha)$, i.e. $P\left(|\phi_{t+1} - \phi_t| \geq \phi_t/(2\alpha)\right) \geq 1/(2\alpha)$. Hence in this case

$$\tau(\epsilon) = O\left(\frac{\ln^2(B)}{\ln^2(1 + \frac{1}{\alpha})\frac{1}{2\alpha}} \ln \epsilon^{-1}\right) = O(\alpha^3 (h + \ln w)^2 \ln \epsilon^{-1}) = O(w^3 (h + \ln w)^2 \ln \epsilon^{-1}).$$

□

3.4.1.2 More complex regions and higher dimensions

In this section, we extend the previous results to higher dimensions and more general regions. To do so, we need to be a bit more formal, so we begin with several definitions.

Given a simply-connected region $\widehat{R}$ in $\mathbb{R}^d$ that is a union of unit cubes on the integer lattice, we associate to it a point set $R = \widehat{R} \cap \mathbb{Z}^d$. In this case we say that R is simply-connected. In three dimensions, a monotonic surface in R is the union of two-dimensional faces such that any cross-section along an axis-aligned plane is a two-dimensional monotonic surface. Such a surface is illustrated in Figure 1b when R is a $2 \times 2 \times 2$ region. In general, given a hypercubic region $R \subset \mathbb{Z}^d$ composed of unit hyper-cubes, a d-dimensional monotonic surface is a set of $(d-1)$-dimensional faces, such that any cross-section along an axis-aligned $(d-1)$-dimensional hyper-plane is a $(d-1)$-dimensional monotonic surface.

We will restrict our focus to a family of simply-connected regions that have favorable properties for the purposes of sampling monotonic surfaces. In order to define this family, we need a few preliminaries.

Definition 3.4.1. *Let $\overline{u}^* = (1, 1, ..., 1) \in \mathbb{Z}^d$. For $v \in \mathbb{Z}^d$, we define $\widehat{r}(v) = \{v + k\overline{u}^* : k \in \mathbb{R}\}$, and we define the* ray $r(v)$ *to be the set*

$$r(v) = \widehat{r}(v) \bigcap \mathbb{Z}^d.$$

Definition 3.4.2. *A d-dimensional simply-connected region $\widehat{R} \subset \mathbb{R}^d$ is* nice *if, for all $v \in \hat{R}$, $\widehat{R} \bigcap \hat{r}(v)$ is connected. We call its associated point set R a* nice *region.*

A nice region $\widehat{R}$ has no holes, and in particular all monotonic surfaces in $\widehat{R}$ are the upper boundary of a subset of cubes in $\widehat{R}$. Note that all hyper-rectangular regions are nice.

Let $\overline{u}_i$ be the unit vector in the ith direction. Given a nice region R, we let $R_L = \{v \in R \text{ such that } \ v - \overline{u}^* \notin R\}$ be the lower envelope of the region.

Definition 3.4.3. *Let $R \subset \mathbb{Z}^d$ be a nice region. A* downset *is a subset $\sigma \subseteq R$, with $R_L \subseteq \sigma$, such that for any i, if $v \in \sigma$ and $v - \overline{u}_i \in R$, then $v - \overline{u}_i \in \sigma$.*

For a nice region R, we define the state space Ω_{mon} to be the set of all downsets of R. We will represent a downset by its *upper boundary*:

Definition 3.4.4. *Let R be any nice region and let σ be any downset of R. We say the* upper boundary *of σ is $\partial(\sigma) = \{v \in \sigma \text{ such that } v + \overline{u}^* \notin \sigma\}$.*

The upper boundary is a monotonic surface in bijection with the downset that defines it. It is important to notice that for any downset σ and point $v \notin \sigma$, if $\sigma \cup v$ is a valid downset, then $|\partial(\sigma)| = |\partial(\sigma \cup v)|$. This is because $\partial(\sigma \cup \{v\}) = \partial(\sigma) \cup \{v\} \setminus \{v - \overline{u}^*\}$. It follows that for any nice region R, the size of the boundary of a valid downset is fixed. This observation motivates the following two definitions that will be convenient when we state the mixing time of our Markov chain.

Definition 3.4.5. *The* span *of a nice region R is $\alpha = |\partial(\sigma)|$, for any downset σ of R.*

Definition 3.4.6. *Let R be any nice region. The* stretch *of R is $\gamma = \max_{v \in R} |R \cap r(v)|$.*

Thus, the stretch is the maximal distance between two points in R in the $\overline{u}^*$ direction. Suppose, for example, that R is an $h \times \cdots \times h$ region in $\mathbb{Z}^d$. Then the span is $\alpha = dh^{d-1}$ and the stretch is $\gamma = h$.

The biased Markov chain

Start at an arbitrary downset, e.g., let $\sigma_0 = R_L$, where R_L is the empty downset, and repeat the following steps. If we are at a downset σ_t at time t, pick a point $v \in \partial(\sigma)$ and an integer $b \in \pm 1$ uniformly at random. If $b = +1$, let $\sigma_{t+1} = \sigma_t \cup (v + \overline{u}^*)$ if this is a valid downset. If $b = -1$, let $\sigma_{t+1} = \sigma_t \setminus \{v\}$ if this is a valid downset. In all other cases, keep σ_t unchanged so that $\sigma_{t+1} = \sigma_t$.

Lemma 3.4.1. *For any nice region R, the Markov chain $\widehat{\mathcal{M}}_{mon}$ connects the state space Ω_{mon}.*

Proof. Let σ be any downset and let v^{max} be any point in σ such that $\sum_i v_i^{max}$ is maximized. We can always remove v^{max} and move to $\sigma' = \sigma \setminus v^{max}$ without violating

the downset condition. Thus, from any valid downset σ we can always remove points and get to the "lowest" downset R_L. Also, such a sequence of steps can be reversed to move from R_L to any other downset ρ. □

Since we have shown that the moves of $\widehat{\mathcal{M}}_{mon}$ connect the state space and all valid moves have the same transition probabilities, we can conclude from detailed balance (Definition 2.1.1) that the chain converges to the uniform distribution over downsets Ω_{mon}. We now define the *biased* Markov chain by using Metropolis-Hastings probabilities [56] to modify the transition probabilities so that we converge to the desired distribution on biased surfaces. This new chain connects the state space by the same argument as in Lemma 3.4.1.

The Biased Markov chain $\mathcal{M}_{mon}$

`Starting at any` σ_0`, iterate the following:`

- `Choose` (v, b, p) `uniformly at random from` $\partial(\sigma_t) \times \{+1, -1\} \times (0, 1)$.
- `If` $b = +1$`, let` $\sigma_{t+1} = \sigma \cup \{v + \overline{u}^*\}$ `if it is a valid downset.`
- `If` $b = -1$ `and` $p \leq \frac{1}{\lambda_v}$`, let` $\sigma_{t+1} = \sigma \setminus \{v\}$ `if it is a valid downset.`
- `Otherwise let` $\sigma_{t+1} = \sigma_t$.

The biased Markov chain $\mathcal{M}_{mon}$ converges to the correct distribution on Ω_{mon} by the detailed balance condition. Moreover, notice that for any given v there is at least one choice of b that proposes a move (adding or removing) that does not result in a valid downset. Therefore $\mathbb{P}[\sigma_{t+1} = \sigma_t] \geq 1/2$, so $\mathcal{M}_{mon}$ is a lazy chain.

Main exponential metric theorem

Next we see how to use the exponential distance metric to bound the mixing time of $\mathcal{M}_{mon}$, showing:

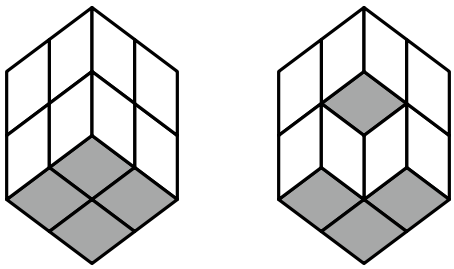


Figure 6: A pair of downsets σ_t (left) and ρ_t (right) where $\rho_t = \sigma_t \cup \{(0,0,0)\}$. These downsets differ on $x = (0,0,0)$, where $\mathcal{M}_{mon}$ can increase ϕ_t by adding $x + \overline{u}_i$, for any i.

Theorem 3.4.2. *Let R be any nice d-dimensional region with volume n, span α, stretch γ, and uniform bias λ. Let $\widehat{\lambda} = (d + \sqrt{d^2 - 4})/2$. If $\sqrt{\lambda} \geq \widehat{\lambda} + \chi$ for some $\chi \geq 0$, then*

1. *If $\chi > 0$, then the mixing time of $\mathcal{M}_{mon}$ satisfies $\tau(\varepsilon) = O\left(\chi^{-2}\alpha(\gamma + \ln n)\ln \varepsilon^{-1}\right)$.*
2. *If $\chi \geq 0$, then $\tau(\varepsilon) = O\left(\alpha^3(\gamma + \ln n)^2\right)\ln(\varepsilon^{-1}))$.*

In particular, if $\lambda \geq d^2$ then $\chi \geq 1$, so $\tau(\varepsilon) = O\left(\alpha(\gamma + \ln n)\ln \varepsilon^{-1}\right)$.

Note that for all nice regions, $\alpha, \gamma < n$, so the mixing time of $\mathcal{M}_{mon}$ is always polynomially bounded for the given biases. When R is an $h_1 \times h_2 \times \cdots \times h_d$ hyperrectangular region, we get optimal bounds as long as each of the h_i are fairly close. Let $h_{min} = \min_i h_i$. The span of R is $\alpha = O(n/h_{min})$ and the stretch is $\gamma = O(h_{min})$. Hence we get $\tau(\varepsilon) = O(n + \frac{n \ln n}{h_{min}} \ln \varepsilon^{-1})$, which is optimal as long as $h_{min} = \Omega(\ln n)$. In particular, we get optimal bounds when the region is a hypercube. Recall that in this case $\alpha = O(h^{d-1})$ and $\gamma = O(h)$, proving Theorem 3.1.2.

Next we define the exponential metric we need to prove Theorem 3.4.2. Let x_0 be any vector in R with maximal L_1 norm. Then define

$$\phi(\sigma, \rho) = \sum_{x \in \sigma \oplus \rho} (\sqrt{\lambda})^{\|x_0\|_1 - \|x\|_1}.$$

We return to the coupling of (σ_t, ρ_t) that simply supplies the same (v^*, b, p) to both σ_t and ρ_t. We let U be the set of downsets that differ on a single cube. However,

instead of the Hamming distance, we use the distance metric

$$\phi(\sigma, \rho) = \sum_{x \in \sigma \oplus \rho} (\sqrt{\lambda})^{\|x_0\|_1 - \|x\|_1},$$

where x_0 is any vector in R with maximal L_1 norm. We will show that this distance metric satisfies non-negative contraction in ϕ_t, which is one of the requirements for Theorem 3.3.1. However, before we can prove that the distances decrease on average, we examine the moves which can *increase* the distance.

For a pair $(\sigma_t, \rho_t) \in U$, there are two different ways the distance can increase in $(\sigma_{t+1}, \rho_{t+1})$. If $\sigma_t = \rho_t \cup \{x\}$, we can increase the distance by attempting to add a v that succeeds in σ_t but fails in ρ_t. This occurs when $v = x + \overline{u}_i$ for some i, so v is "supported" in σ_t but not ρ_t. The other way to increase the distance between σ_t and ρ_t is to remove a v that succeeds in ρ_t but not in σ_t. This occurs when $v = x - \overline{u}_i$ for some i, as the move creates a valid downset in ρ but not in σ. The following lemma bounds the number of such increases in distance.

Lemma 3.4.3. *For $\sigma_t = \rho_t \cup \{x\}$, there are at most d choices of (v^*, b) such that ϕ_t increases.*

Proof. We prove the Lemma by claiming that for dimensions $i \neq j$, if $\mathcal{M}_{mon}$ can increase the distance by choosing $v = x + \overline{u}_i$, then it cannot increase the distance by choosing $v = x - \overline{u}_j$. This follows from a proof by contradiction: If $\mathcal{M}_{mon}$ can increase the distance with $x + \overline{u}_i$, then it is because $\rho_{t+1} = \rho_t \cup \{x + \overline{u}_i\}$ is a valid downset. That means $x + \overline{u}_i - \overline{u}_j \in \rho_t$. On the other hand, if $\mathcal{M}_{mon}$ can increase the distance with $x - \overline{u}_j$, it is because $\sigma_{t+1} = \sigma_t \backslash \{x - \overline{u}_j\}$ is a valid downset, which is only true if $x - \overline{u}_j + \overline{u}_i \notin \sigma_t$. But this contradicts the fact $\sigma_t \oplus \rho_t = \{x\}$, justifying our claim.

This implies that, to increase the distance, $\mathcal{M}_{mon}$ may add vectors of the form $x + \overline{u}_i$ for various dimensions i as in Figure 6, or it may remove vectors of the form $x - \overline{u}_i$ for various dimensions i, or it may add $x + \overline{u}_i$ and remove $x - \overline{u}_i$ in a single

dimension i, as in Figure 7. In each of these cases, there are at most d choices of v that increase the distance. □

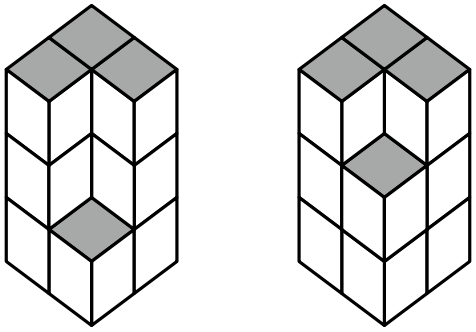


Figure 7: Downsets that differ on x, where $\mathcal{M}_{mon}$ increases ϕ_t by adding the vector above x or removing the vector below x.

We may now prove Theorem 3.4.2.

Proof. We will show that the distance metric ϕ defined above satisfies the conditions of Path Coupling Theorem for Exponential metrics, Theorem 3.3.1. First we want to show that the expected change in ϕ is negative. By Lemma 3.4.3, there are at most d choices of (v^*, b) that can increase ϕ_t. We claim that each of these has an expected increase of at most $\phi_t \lambda^{-1/2}$. To see this, consider a move of the form $v = x + \overline{u}_i$ for some i. Then the increase in distance is $(\sqrt{\lambda})^{\|x_0\|_1 - \|v\|_1} = \phi_t \lambda^{-1/2}$. If the move is of the form $v = x - \overline{u}_i$ for some i, then the increase in distance is $\lambda^{-\|v\|_1/2} = \phi_t \sqrt{\lambda}$, but the chance of choosing an appropriate p is $1/\lambda$. Therefore the expected increase is at most $\phi_t \lambda^{-1/2}$.

There are also two choices of (v^*, b) that *decrease* ϕ_t; corresponding to adding x and removing x. These each decrease ϕ_t by ϕ_t, and succeed with probability 1 and

$1/\lambda$, respectively. Therefore the expected change in distance satisfies

$$\begin{aligned}\mathbb{E}[\phi_{t+1} - \phi_t] &\leq \frac{1}{2\alpha}\left(d \cdot \frac{\phi_t}{\sqrt{\lambda}} - \left(1 + \frac{1}{\lambda}\right)\phi_t\right) \\ &= \frac{\phi_t}{2\alpha}\left(\frac{d}{\sqrt{\lambda}} - 1 - \frac{1}{\lambda}\right) \\ &\leq -\frac{\phi_t \chi}{2\alpha}.\end{aligned}$$

Next we check the other conditions of Theorem 3.3.1. For arbitrary $\sigma, \rho \in \Omega_{mon}$, if $x \in \sigma \oplus \rho$ for some x, then $\phi(\sigma, \rho) \geq \sqrt{\lambda}^{\|x_0\|_1 - \|x\|_1} \geq 1$. Therefore if $\phi(\sigma, \rho) < 1$, $\phi(\sigma, \rho) = 0$. Let U be the set of pairs of downsets that differ on a single vector. For arbitrary $\sigma, \rho \in \Omega_{mon}$, we can connect σ to ρ by simply adding or removing the vectors in $\sigma \oplus \rho$ one by one, and $\phi(\sigma, \rho)$ is the sum of the distances. Since the volume of $\widehat{R}$ is n, there are at most n possible vectors in $\sigma \oplus \rho$, so $\phi(\sigma, \rho) \leq n\lambda^{\gamma/2}$ for all σ, ρ.

We consider two cases. If $\lambda - \widehat{\lambda}_d > \chi$, then $\mathbb{E}[\phi_{t+1}] \leq \beta\phi_t$, where $\beta = \Theta(1/\alpha)$. Thus, by Theorem 3.3.1, we have $\tau(\epsilon) = O(\chi^{-2}\alpha(\gamma + \ln n)\ln \epsilon^{-1})$. On the other hand, if $\lambda - \widehat{\lambda}_d > 0$, but less than any constant, then we use the second part of Theorem 3.3.1. For any pair of σ, ρ, $\mathcal{M}_{mon}$ can always add a vector v^* in their difference that maximizes $\|v\|_1$. This would change ϕ_t by at least ϕ_t/α. The appropriate v^* is chosen with probability at least $1/\alpha$ and the appropriate b is chosen with probability $1/2$ (and every p succeeds when adding). Therefore there is a $1/(2\alpha)$ chance of changing ϕ_t by $\phi_t/(2\alpha)$, i.e. $P\left(|\phi_{t+1} - \phi_t| \geq \phi_t/(2\alpha)\right) \geq 1/(2\alpha)$. Hence in this case

$$\tau(\epsilon) = O\left(\frac{\ln^2(B)}{\ln^2(1 + \frac{1}{\alpha})\frac{1}{2\alpha}} \ln \epsilon^{-1}\right) = O(\alpha^3(\gamma + \ln n)^2 \ln \epsilon^{-1}).$$

□

3.4.2 Hitting time to the maximal tiling

We now introduce a second technique that allows us to get improved bounds for the mixing rate of the uniform bias Markov chain $\mathcal{M}_{mon}$ whenever $\lambda \geq d$ and the region is an $h \times h \times \cdots h$ hypercube of volume $n = h^d$. Specifically, we prove:

Theorem 3.4.4. *Let $R \subset \mathbb{Z}^d$ be the $h \times h \times \cdots \times h$ hypercube of volume $n = h^d$ and bias $\lambda \geq d$. Then the mixing time of $\mathcal{M}_{mon}$ satisfies $\tau(\varepsilon) = O(h^{2d-1} \ln \varepsilon^{-1})$. In general, this is $o(n^2 \ln \varepsilon^{-1})$, or in 2 dimensions, $O(n\sqrt{n} \ln \varepsilon^{-1})$.*

The proof relies on the monotonicity of $\mathcal{M}_{mon}$ with respect to the trivial coupling. In other words, if (X_t, Y_t) are coupled and $X_t \subseteq Y_t$, then after one step of the coupling, $X_{t+1} \subseteq Y_{t+1}$. This implies that the coupling time is bounded by the time to hit the full cube F starting from the empty cube F', and we can show that this will happen quickly because the distance to F is always non-increasing in expectation (see Fig. 8b).

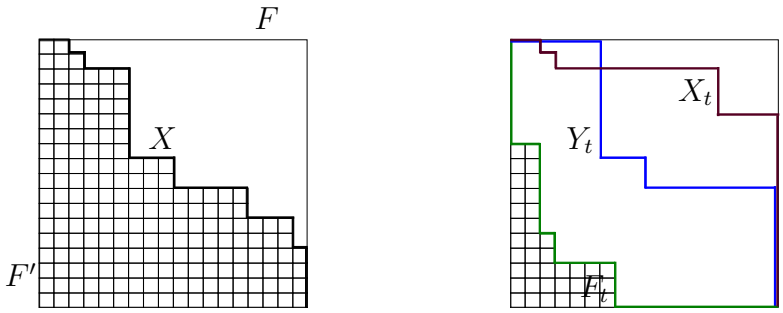


Figure 8: (a) A staircase walk with 5 peaks and 4 valleys. (b) $F_t \subseteq X_t$ and $F_t \subseteq Y_t$ for all t

We begin by proving that a monotone Markov chain with nonnegative drift will mix rapidly. Call a Markov chain $\mathcal{M}$ on Ω *monotone* if Ω is a distributive lattice with partial order $\preccurlyeq$ and there exists a coupling (X_t, Y_t) such that if $X_t \preccurlyeq Y_t$, then $X_{t+1} \preccurlyeq Y_{t+1}$. For $X_t \in \Omega$, let $\mathbb{P}_{\prec}(X_t) = \mathbb{P}[X_{t+1} \neq X_t \text{ and } X_{t+1} \preccurlyeq X_t]$ and let $\mathbb{P}_{\succ}(X_t) = \mathbb{P}[X_{t+1} \neq X_t \text{ and } X_t \preccurlyeq X_{t+1}]$.

Theorem 3.4.5. *Given a monotone Markov chain $\mathcal{M}$ on Ω with maximal element F and minimal element F', let $h(X, Y)$ be the Hamming distance between X and Y and let $H = \max_{X,Y \in \Omega}\{h(X, Y)\}$. Assume that for all $X \neq F \in \Omega$,*

$$\mathbb{P}_{\succ}(X) - \mathbb{P}_{\prec}(X) \geq \kappa \geq 0.$$

1. *If $\kappa > 0$ then the mixing time is $\tau(\epsilon) \leq \lceil \frac{eH \ln H}{\kappa} \rceil \lceil \ln(\epsilon^{-1}) \rceil$.*

2. *If $\kappa = 0$ and for all $t \geq 0$, $\mathbb{P}[X' \neq X] \geq Q$ then the mixing time satisfies $\tau(\epsilon) \leq \lceil \frac{2eH^2}{Q} \rceil \lceil \ln(\epsilon^{-1}) \rceil$.*

Proof. First we notice that the coupling time $T^{x,y} = \mathbb{E}[\min\{t : X_t = Y_t | X_0 = x, Y_0 = y\}]$ is bounded by the hitting time to reach F from F'. Define $T_F^x = \mathbb{E}[\min\{t : X_t = F | X_0 = x\}]$ to be the hitting time to reach F from x. Let $F_0 = F'$, and couple the moves of $\{F_t\}$ together with the moves of $\{X_t\}$ and $\{Y_t\}$; that is, choose the same (i, b, r) for F_t as in X_t and Y_t. Since the Markov chain is monotone with respect to the given coupling, we have $F_t \preccurlyeq X_t$ and $F_t \preccurlyeq Y_t$ for all $t \geq 0$. Thus if $F_t = F$, we also have $X_t = Y_t = F$. So the coupling time for X_t and Y_t is bounded by the hitting time of F_t to F. See Figure 8(b).

Let $\phi_t = h(X_t, F)$. For part (1), we use $\mathbb{E}[\phi_t] \leq \phi_t - \kappa\phi_{t-1}/H \leq (1 - \kappa/H)^t \phi_0$. For part (2), note that $\mathbb{E}[\phi_{t+1}^2 + \phi_t^2] \geq \phi_t^2 \geq 1$, so using Lemma 3.3.2 with $q = 0$ and $D = H$, we obtain $\mathbb{E}[T] \leq \frac{2H^2}{Q}$. Then the theorem follows from the Coupling Lemma, Theorem 2.2.4. □

Notice that since $\lambda > 1$, we expect that for any downset $X_0 \in \Omega$, the sequence $\{X_t\}$ should approach the unique maximal element of Ω. We will show that it suffices to choose $\lambda \geq d$ to reach the full $d-$dimensional cube from an arbitrary position in polynomial time, thus achieving polynomial mixing time. In particular, we will show that $\mathcal{M}_{mon}$ satisfies the conditions of Theorem 3.4.5.

We define a *peak* of X to be a position where we can remove a hypercube from X and a *valley* of X to be a position where we can add a hypercube to X (see Figure 8(a)). For the following lemma, define $\mathcal{V}(D)$(and, respectively, $\mathcal{P}(D)$) to be the set of valleys (peaks) of a downset D.

Lemma 3.4.6. *For any downset $\sigma \neq F \in \Omega$,*

$$|\mathcal{P}(\sigma)| \leq (d-1)|\mathcal{V}(\sigma)| + 1. \tag{7}$$

Proof of Lemma 3.4.6. First we notice that if $\sigma \neq F$, then it has at least one valley. Furthermore, if σ has a single valley, then the number of peaks is at most $d = (d-1)\mathcal{V}(\sigma) + 1$. Now, assume the number of valleys is more than one and proceed as follows. Choose a valley $v = (v_1, v_2, \ldots, v_d)$ that maximizes v_d. Construct σ' from σ by adding every hypercube lying above v in the d dimension; that is, $\sigma' = \sigma \cup \{v + e_d, v + 2e_d, \ldots, v + (h - v_d)e_d\}$. Then $|\mathcal{P}(\sigma')| \geq |\mathcal{P}(\sigma)| - d + 1$, and $|\mathcal{V}(\sigma')| = |\mathcal{V}(\sigma)| - 1$. Hence

$$|\mathcal{P}(\sigma)| - (d-1)|\mathcal{V}(\sigma)| \leq |\mathcal{P}(\sigma')| - (d-1)(|\mathcal{V}(\sigma')|.$$

Obtain $\widehat{\sigma}$ by iterating this operation until only a single valley remains; then we have

$$|\mathcal{P}(\sigma)| - (d-1)|\mathcal{V}(\sigma)| \leq |\mathcal{P}(\widehat{\sigma})| - (d-1)|\mathcal{V}(\widehat{\sigma})| \leq 1,$$

as desired. □

Lemma 3.4.6 implies that if $\lambda \geq d$ then for all $\sigma \neq F \in \Omega$, $|\mathcal{P}(\sigma)| \leq (d-1)|\mathcal{V}(\sigma)| + 1 \leq d|\mathcal{V}(\sigma)|$, and so

$$\begin{aligned} \mathbb{P}_{\succ}(\sigma) - \mathbb{P}_{\prec}(\sigma) &= \frac{1}{2\alpha}\left(|\mathcal{V}(\sigma)| - \frac{1}{\lambda}|\mathcal{P}(\sigma)|\right) \\ &\geq \frac{|\mathcal{V}(\sigma)|}{2\alpha}\left(1 - \frac{d}{\lambda}\right) \geq 0. \end{aligned}$$

Hence, the Markov chain $\mathcal{M}_{mon}$ has nonnegative drift towards the maximal configuration. Using Theorem 3.4.5 we can show that for $\lambda = d + \delta$ for some $\delta \geq 0$, $\tau(\varepsilon) = O(\delta^{-1}\alpha n \ln n \ln \varepsilon^{-1})$ for $\delta > 0$, and otherwise $\tau(\varepsilon) = O(\alpha n^2 \ln \varepsilon^{-1})$. However, we are able to get a better bound in Theorem 3.4.4 by introducing another metric that gives strict additive contraction for every configuration, even when $\lambda = d$.

Proof of Theorem 3.4.4.

Recall from the proof of Theorem 3.4.5 that the coupling time is bounded by the hitting time to reach F from F'. We will show that the expected time to hit

F is small, using Lemma 3.4.6. Let $S_1 = \{\sigma \in \Omega : \mathcal{P}(\sigma) < d\mathcal{V}(\sigma)\} \cup F$, and $S_2 = \{\sigma \in \Omega : \mathcal{P}(\sigma) = d\mathcal{V}(\sigma)\}$. Define the function $\phi(\sigma) = H(F, \sigma) + I_{S_2}/(2d)$, where $H(X, Y)$ is the Hamming distance between the downsets X and Y, and I_{S_2} is the indicator function for the set S_2. We will show that ϕ has negative additive drift towards 0.

Notice that Lemma 3.4.6 implies that if $\sigma \in S_2$ then it has exactly one valley. Moreover, if $\sigma \in S_1$ can move to S_2 in a single step, then the number of possible moves to take it to S_2 is at most 3. If σ has a single valley then it must have a valley v such that for some dimension i, $v_i = 0$; adding a cube at this valley could move σ into S_2. If σ has three valleys then it has a single hypercube c it can remove to enter S_2, and if σ has two valleys then it can enter S_2 by adding cubes at either of those valleys or removing a cube between them are the only ways to potentially move into S_2. Hence, if $\sigma_t \in S_1$, then the probability that $\sigma_{t+1} \in S_2$ is at most $3/(2\alpha)$. Moreover, if $\sigma_t \in S_2$, then the probability that $\sigma_{t+1} \in S_1$ is at least $1/(2\alpha)$.

Now, conditioning on whether σ_t is in S_1 or in S_2, we have $\mathbb{E}[\phi(\sigma_{t+1}) - \phi(\sigma_t)|\sigma_t \in S_2] \leq 0 - 2d)^{-1}(2\alpha)^{-1}$ and

$$\begin{aligned}
\mathbb{E}[\phi(\sigma_{t+1}) - \phi(\sigma_t)|\sigma_t \in S_1] &\leq \frac{-|\mathcal{V}(\sigma_t)| + \lambda^{-1}|\mathcal{P}(\sigma_t)|}{2\alpha} + \frac{3}{2\alpha}\left(\frac{1}{2d}\right) \\
&\leq \frac{1}{2\alpha}\left(-|\mathcal{V}(\sigma_t)| + \lambda^{-1}(d-1)|\mathcal{V}(\sigma_t)|\right) + \frac{3}{2\alpha}\left(\frac{1}{2d}\right) \\
&\leq \frac{|\mathcal{V}(\sigma_t)|}{2\alpha}\left(-\frac{1}{d}\right) + \frac{3}{2\alpha}\left(\frac{1}{2d}\right) \\
&\leq \frac{-1}{4\alpha d}.
\end{aligned}$$

Thus

$$\mathbb{E}[\phi(\sigma_t)] \leq \sigma_0 - \frac{t}{4\alpha d} \leq \left(n + \frac{1}{2d}\right) - \frac{t}{4\alpha d} \leq \epsilon$$

whenever $t \geq 4d\alpha(n + \frac{1}{2d} - \epsilon)$. Thus the mixing time satisfies $\tau(\varepsilon)O(\alpha n \ln(\varepsilon^{-1})) = O(h^{2d-1}\ln(\varepsilon^{-1})) = O(n^2 \ln(\varepsilon^{-1}))$ for arbitrary dimension d, but for $d = 2$, this implies $\tau(\varepsilon) = O(h^3 \ln(\varepsilon^{-1})) = O(n\sqrt{n}\ln(\varepsilon^{-1}))$.

□

3.5 The fluctuating bias Markov chain

Consider the more general setting, where the bias of a tile depends on its position. This situation is more realistic, particularly in the context of self-assembly. For instance, the probability of a tile lined with DNA attaching to the substrate depends on the strength of the bonds along the edges of the tile as well as the relative densities of each tile. Recall the *bias at* $\bar{x}$ is defined as follows: if τ is formed by adding a cube at position $\bar{x}$ to σ, then $\lambda_{\bar{x}} = P(\sigma, \tau)/P(\tau, \sigma)$ is called the *bias at* $\bar{x}$. The stationary probability of a configuration σ will be proportional to $\prod_{\bar{x} \in \sigma} \lambda_{\bar{x}}$, where by $\bar{x} \in \sigma$ we mean that $\bar{x}$ is a cube lying below the surface σ.

It turns out that most of the results from Section 3.4 generalize to this setting, as long as we satisfy certain bounds on the amount the bias can fluctuate. When the minimum bias λ_L is large enough, we do not need any upper bound. While the upper bound might seem unnecessary even for small values of $\lambda_L > 1$, the chain can actually take exponential time to reach equilibrium if the biases vary too much (see Section 3.5.3).

3.5.1 Fast mixing with large enough minimum bias

In Section 3.4.2, we showed that: (i) the coupling time for the Markov chain $\mathcal{M}_{mon}$ was bounded by the hitting time to the maximal configuration, and (ii) the hitting time is polynomial, assuming that the bias is large enough. Clearly, (i) still holds for the fluctuating bias Markov chain $\mathcal{M}_{til}^{fluct}$, and the hitting time to the maximal configuration should only decrease if some of the biases are increased. Therefore, Theorem 3.1.3 is a simple consequence of Theorem 3.4.4.

Moreover, we can use a similar argument, together with the uniform bias results above, to obtain the stronger result for fluctuating bias in 2 dimensions given in Theorem 3.1.4.

Proof of Theorem 3.1.4. It is well-known that in 2 dimensions for uniform bias λ, where $\lambda > 1$ is a constant, the maximal configuration has constant probability in the stationary distribution (see, for example, Benjamini et al. [5]). By Theorem 3.1.1, we know that the mixing time of the uniform bias chain with bias λ_L is $O\left(w(h + \ln w) \ln \varepsilon^{-1}\right)$, so we expect the uniform bias chain to hit the maximal configuration in $O\left(w(h + \ln w) \ln \varepsilon^{-1}\right)$ steps. Since the fluctuating bias Markov chain always lies above the uniform bias chain, it must also hit the maximal configuration in that expected time. Recalling from Theorem 3.4.5 that the hitting time to the maximal configuration is an upper bound on the coupling time, this proves the theorem. □

3.5.2 Fast mixing when the fluctuations are bounded

We can also extend the exponential metric technique of Section 3.4.1 to handle fluctuating bias, provided the biases λ_x for $x \in R$ do not vary too much.

Theorem 3.5.1. *Let R be any nice d-dimensional region with volume n, span α, stretch γ, and suppose the bias at any point x satisfies $1 < \lambda_L \leq \lambda_x \leq \lambda_U$. If the maximum and minimum biases are such that*

$$\frac{d}{\sqrt{\lambda_L}} - 1 - \frac{1}{\lambda_U} \leq -\chi \tag{8}$$

for some $\chi \geq 0$, then

1. *If $\chi > 0$, then the mixing time of $\mathcal{M}_{til}^{fluct}$ satisfies $\tau(\varepsilon) = O\left(\chi^{-1}\alpha(\gamma + \ln n) \ln \varepsilon^{-1}\right)$.*
2. *If $\chi \geq 0$, then $\tau(\varepsilon) = O\left(\alpha^3(\gamma + \ln n)^2) \ln(\varepsilon^{-1})\right)$.*

Proof. This theorem is proved nearly identically to Theorem 3.4.2. In this case we define the distance metric as

$$\phi(\sigma, \rho) = \sum_{x \in \sigma \oplus \rho} (\sqrt{\lambda_L})^{\|x_0\|_1 - \|x\|_1}.$$

Given that $\mathcal{M}_{til}^{fluct}$ chose v^* and b such that ϕ_t can increase, the expected increase is at most $\phi_t {\lambda_L}^{-1/2}$. Indeed, if the move is of the form $v = x + \overline{u}_i$ for some i, then the increase in distance is $(\sqrt{\lambda_L})^{\|x_0\|_1 - \|v\|_1} = \phi_t {\lambda_L}^{-1/2}$. If the move is of the form $v = x - \overline{u}_i$ for some i, then the increase in distance is $\lambda_L^{-\|v\|_1/2} = \phi_t \sqrt{\lambda_L}$, but the chance of choosing an appropriate p is $1/\lambda_v \leq 1/\lambda_L$. Therefore the expected increase is again at most $\phi_t {\lambda_L}^{-1/2}$.

This implies that the expected change in distance is negative. As before, there are at most d bad moves, but now the two good moves happen with probability 1 and $1/\lambda_x \geq 1/\lambda_U$, respectively. Therefore the expected change in distance satisfies

$$\mathbb{E}[\phi_{t+1} - \phi_t] \leq \frac{1}{2\alpha}\left(d \cdot \frac{\phi_t}{\sqrt{\lambda_L}} - \left(1 + \frac{1}{\lambda_U}\right)\phi_t\right) \;\leq\; -\frac{\phi_t \chi}{2\alpha}.$$

The rest of the proof is identical to the proof of Theorem 3.4.2. □

3.5.3 Slow mixing when the fluctuations are unbounded

Note that we have restricted throughout this chapter to the case when $\lambda_{x,y} \geq 1$ for all (x, y). This restriction is necessary, or the chain might not be rapidly mixing. For example, it can be shown that if $\lambda_{x,y} < 1$ when $x+y \leq n$ and $\lambda_{x,y} > 1$ when $x+y > n$, then $\mathcal{M}_{til}^{fluct}$ requires exponential time to converge. Indeed it will be difficult to move from a tiling that is nearly empty to one that is nearly full, even though these each occupy a constant fraction of the stationary probability. In fact, we will see presently that even if $\lambda_L > 1$, the Markov chain may have exponential mixing time.

Let R be a square region in 2 dimensions. We consider the following question. Suppose that for all $(x, y) \in R$, $\lambda_{x,y} \geq \lambda_L > 1$. We know by Theorem 3.4.2 that for uniform bias, the Markov chain is rapidly mixing for all $\lambda > 1$ polynomially bounded away from 1 and the result is in general easier to show for larger values of λ. This leads us to expect rapid mixing in the fluctuating bias case, as long as the minimum bias λ_L is polynomially bounded away from 1. However, this is not true in general. We show that if the biases below the line $x + y = n + M$ (where $M = 2n^{2/3}$) are all

close to 1 and all other biases are some very large constant ξ, then the mixing time of $\mathcal{M}_{mon}$ is exponentially large in n (see Figure 9). We identify sets S_1, S_2, S_3 such that $\pi(S_2)$ is exponentially smaller than both $\pi(S_1)$ and $\pi(S_3)$, but to get between S_1 and S_3, $\mathcal{M}_{nn}$ and $\mathcal{M}_T$ must pass through S_2, the cut. Then we use the conductance to prove $\mathcal{M}_{til}^{fluct}$ is slowly mixing.

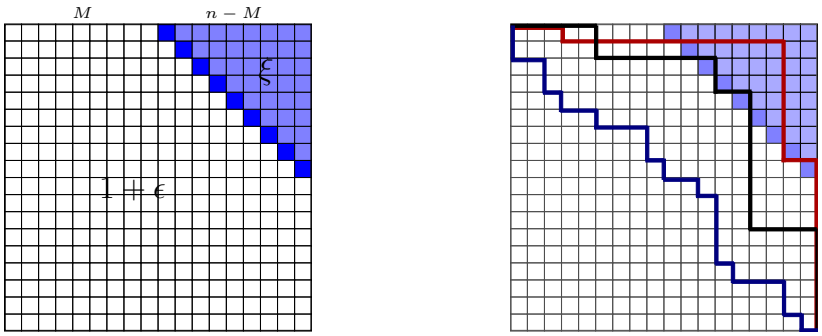


Figure 9: (a) Fluctuating bias with exponential mixing time. (b) Staircase walks in S_1, S_2, and S_3.

Proof of Theorem 3.1.5.

Proof. We will define values $\{\lambda_{x,y}\}_{(x,y)\in R}$ such that the conductance of the Markov chain $\mathcal{M}_{til}^{fluct}$ is small. Let $M = 2n^{2/3}$. For all (x, y) such that $x + y \leq n + M$, define $\lambda_{x,y} = \epsilon = 1 + \frac{c_1}{n^2}$, where $c_1 \in \mathbb{R}$ is a constant. For all remaining (x, y), let $\lambda_{x,y} = \xi$, where $\xi > 1$ is a constant. For a staircase walk σ, define the *height of* σ_i as $\sum_{j\leq i} \sigma_j$, and let $\max(\sigma)$ be the maximum height of σ_i over all $1 \leq i \leq 2n$. Let S_1 be the set of configurations σ such that $\max(\sigma) < n + M$, S_2 the set of configurations such that $\max(\sigma) = n + M$, and S_3 the set of configurations such that $\max(\sigma) > n + M$.

First we notice that since the maximal tiling is in S_3,

$$\Pi(S_3) \geq \frac{1}{Z}(1+\epsilon)^{n^2-\frac{(n-M)^2}{2}}(\xi)^{\frac{(n-M)^2}{2}}.$$

Also, $\Pi(S_1) = \frac{1}{Z}\sum_{\sigma\in S_1}(1+\epsilon)^{A(\sigma)}$, where $A(\sigma)$ is the number of tiles contained in σ.

Therefore

$$\begin{aligned}
\Pi(S_1) &\leq \frac{1}{Z}\sum_{\sigma\in S_1}(1+\epsilon)^{n^2-\frac{(n-M)^2}{2}} \\
&\leq \frac{1}{Z}\binom{2n}{n}(1+\epsilon)^{n^2-\frac{(n-M)^2}{2}} \\
&\leq \frac{1}{Z}(2e)^n(1+\epsilon)^{n^2-\frac{(n-M)^2}{2}} \\
&\leq \frac{1}{Z}(1+\epsilon)^{n^2-\frac{(n-M)^2}{2}}(\xi)^{\frac{(n-M)^2}{2}} \;\leq\; \Pi(S_3)
\end{aligned}$$

since ξ is a constant larger than 1. Hence $\Pi(S_1) \leq \Pi(S_3)$. We will show that $\Pi(S_2)$ is exponentially small in comparison to $\Pi(S_1)$ (and hence also to $\Pi(S_3)$).

$$\Pi(S_2) = \frac{1}{Z}\sum_{\sigma\in S_2}(1+\epsilon)^{A(\sigma)} \leq \frac{(1+\epsilon)^{n^2}|S_2|}{Z}.$$

We bound $|S_2|$ as follows. The unbiased Markov chain is equivalent to a simple random walk $W_{2n} = X_1 + X_2 + \cdots + X_{2n} = 0$, where $X_i \in \{+1, -1\}$ and where a $+1$ represents a step to the right and a -1 represents a step down. We call this random walk *tethered* since it is required to end at 0 after $2n$ steps. Compare walk W_{2n} with the untethered simple random walk $W'_{2n} = X'_1 + X'_2 + \ldots + X'_{2n}$.

$$\begin{aligned}
P\left(\max_{1\leq t\leq 2n} W_t \geq M\right) &= P\left(\max_{1\leq t\leq 2n} W'_t \geq M \mid W'_{2n} = 0\right) \\
&= \frac{P\left(\max_{1\leq t\leq 2n} W'_t \geq M\right)}{P(W'_{2n}=0)} \\
&= \frac{2^{2n}}{\binom{2n}{n}} P\left(\max_{1\leq t\leq 2n} W'_t \geq M\right) \\
&\approx \sqrt{\pi n}\;\; P\left(\max_{1\leq t\leq 2n} W'_t \geq M\right).
\end{aligned}$$

Since the $\{X'_i\}$ are independent, we can use Chernoff bounds to see that

$$P\left(\max_{1\leq t\leq 2n} W'_t \geq M\right) \;\leq\; 2nP(W'_{2n} \geq M) \;\leq\; 2ne^{\frac{-M^2}{2n}}.$$

Together these show that $P\left(\max_{1\leq t\leq 2n} W_t \geq M\right) < e^{-n^{1/3}}$, by definition of M. Therefore we have

$$\begin{aligned}
\Pi(S_2) &\leq \frac{1}{Z}(1+\epsilon)^{n^2}|S_2 \cup S_3| \\
&\leq \frac{1}{Z}\binom{2n}{n}e^{-n^{1/3}} \\
&\leq \frac{1}{Z}\binom{2n}{n}e^{-n^{1/3}+1}(1-e^{-n^{1/3}}) \\
&\leq \frac{1}{Z}|S_1|e^{-n^{1/3}+1} \\
&\leq e^{-n^{1/3}+1}\Pi(S_1),
\end{aligned}$$

as desired. Thus, the conductance satisfies

$$\phi \leq \sum_{x\in S_1}\frac{\pi(x)}{\pi(S_1)}\sum_{y\in S_2}P(x,y) \leq \sum_{x\in S_1}\frac{\pi(x)}{\pi(S_1)}\pi(S_2) \leq e^{-n^{1/3}+1}\pi(S_1) \leq e^{-n^{1/3}+1}/2.$$

Hence, by Theorem 2.2.2, the mixing time of $\mathcal{M}_{mon}$ is at least $e^{n^{1/3}-1}/4 - 1/2$. □

3.6 Conclusion

In this chapter we show that the DNA-based self-assembly is efficient for any uniform bias in two dimensions and for large enough bias in higher dimensions. We also give the first analysis of self-assembly with fluctuating bias, showing that as long as the biases do not differ by too much, the assembly is still rapid. We make several contributions to the study of biased card shuffling as well. The bound in Theorem 3.1.1 on the mixing time of $\mathcal{M}_{mon}$ when $d = 2$ yields a simpler proof that the nearest-neighbor transposition chain on biased permutations is rapidly mixing, using the bijection from Benjamini et al. [5]. In fact, we achieve the same optimal bounds on the mixing time. Recently our improved bounds on the mixing time of $\mathcal{M}_{mon}$ for rectangular regions were used to show a tighter bound on the mixing time of the nearest-neighbor transposition chain for a generalization of biased permutations arising in the context of self-organizing lists [6].

The technique from Section 3.4.1 can be extended to other applications, such as biased 3-colorings. There is a well-known bijection between 3-colorings of $\mathbb{Z}^2$ and sets of monotonic, *edge*-disjoint paths (see, e.g. [49]). The construction generalizes to arbitrary dimension as well, forming $(d-1)$-dimensional monotonic surfaces that are *face*-disjoint. There is a Markov chain $\mathcal{M}_{col}$ arising in the context of asynchronous cellular automata which samples biased 3-colorings. Our technique can be used to show that as long as the bias satisfies $\lambda \geq 4d^2$, the mixing time of $\mathcal{M}_{Col}$ satisfies $\tau(\varepsilon) = O\left(n^2 \ln(\varepsilon^{-1})\right)$. We expect that there may be other situations in which this technique could be useful as well.

CHAPTER IV

MIXING TIMES OF SELF-ORGANIZING LISTS AND BIASED PERMUTATIONS

Next, we consider the convergence rate of a Markov chain $\mathcal{M}_{nn}$ for sampling biased permutations that arises in the context of self-organizing lists. We prove that $\mathcal{M}_{nn}$ is always rapidly mixing for two general classes of positively biased $\{p_{xy}\}$. More significantly, we also disprove a general conjecture by exhibiting positively biased probabilities for which the transposition chain will require exponential time to converge.

4.1 Self-organizing lists and biased permutations

Sampling random permutations of n elements is one of the most fundamental problems in probability theory. A natural Markov chain that has been studied extensively is a symmetric chain, $\mathcal{M}_{nn}$, that makes nearest neighbor transpositions. Aldous [1] and Diaconis and Shashahani [19] showed that $\mathcal{M}_{nn}$ converges to the uniform distribution in $\Omega(n^3)$ and $O(n^3 \log n)$ steps. This was improved by Wilson [67] who showed a tight bound of $\Theta(n^3 \log n)$. There has been interest in generalizing this result to cases where the probability of inverting some (or all) pairs is not uniform, but thus far there has been very little progress on the general problem.

The Markov chain $\mathcal{M}_{nn}$ is equivalent to the Move-Ahead-One (or *transpose*) algorithm for self-organizing lists [36]. Fill [27, 28] introduced one biased version of $\mathcal{M}_{nn}$ in this context. We present a variant of his model where we have n records, each associated with a Poisson clock with rate 1. When a clock rings, we try to move the associated record i ahead of its immediate predecessor in the list, if it exists.

If it has higher rank than its predecessor, then it always succeeds, while if its rank is lower we move it ahead with probability $f_i \leq 1$. The rank is determined by an underlying complete ordering on the n records. In Fill's setting, the records are pages in a linked list and the f_i values are determined by the frequency of page requests, so more frequently requested pages are further up on the list and require smaller access time.

More generally, consider a permutation on objects $\{1, \ldots, n\}$ and a set of input parameters $\mathbf{P} = \{p_{x,y}\}$ with $p_{x,y} \leq 1$. At each point in time, the Markov chain picks a pair of adjacent elements in the current permutation, say x and y in positions $i-1$ and i. If $x < y$ then we put x in front of y with probability $p_{x,y}$ and we put y in front of x with probability $1 - p_{x,y}$. We are interested in understanding for which choices of $\mathbf{P}$ the chain is rapidly mixing, or quickly converging to equilibrium. We call a distribution on permutations *positively biased* if it arises from $\mathcal{M}_{nn}$ using such a $\mathbf{P}$ vector where we are at least as likely to put each pair in order as out of order; that is, $p_{x,y} \geq 1/2$ for all $x < y$. It is easy to see that without this restriction, the chain can take exponential time to converge. However, if $p_{x,y} \geq 1/2$ for all $x < y$, then object 1, which has the highest rank, tends to move toward the front of the list, and object n, which has the lowest rank, will tend to move toward the back. Fill asked whether $\mathcal{M}_{nn}$ is always rapidly mixing for positively biased permutations. Specifically, in a widely circulated manuscript [28], Fill conjectured that if $\mathbf{P}$ satisfies a "regularity condition" requiring the $p_{x,y}$ to be concave in x and in y then the spectral gap is maximized when $p_{x,y} = 1/2$ for all x and y. He calls this *the gap problem* for $\mathcal{M}_{nn}$.

There has been very little progress on the gap problem. Fill verified the conjecture for $n \leq 3$ and gives experimental evidence for $n \leq 5$ [28]. In 2006, Benjamini et al. [5] solved the constant bias case by showing that if, for all $x < y$, $p_{x,y} = p$ for some $1/2 < p < 1$, then the chain always mixes in time $\Theta(n^2)$. Their proof relates this constant bias case to an Asymmetric Simple Exclusion Process (ASEP). An ASEP

is a Markov chain on sequences of k_1 zeros and k_2 ones that is more likely to move a one to the left of a zero than the reverse. This reduction from permutations to ASEP does not hold when the biases are not constant, so their proof does not readily extend to the general case. The only other setting of biased permutations in which there are rigorous bounds on the mixing time is when each $p_{x,y}$ is either $1/2$ or 1. In this case, the results of Bubley and Dyer [11] for finding linear extensions of a partial order imply that the chain mixes in time $O(n^3 \log n)$.

Our results: In this paper we prove the conjecture for two new classes of $\mathbf{P}$ values, both of which include the constant bias setting as a special case. For the first class, we are given $r_1, \ldots, r_{n-1}$ and $p_{x,y} = r_x$ if $x < y$, so the probability of an inversion depends only on the object with higher rank (using similar arguments we can also handle the case where the probability depends on the object with lower rank). This case is precisely the self-organizing list application that motivates this work; when a record x is woken up, it checks if its predecessor has higher or lower rank than it. If lower, it moves ahead with probability 1, but if higher, then it moves ahead with probability $f_x = r_x/(1 + r_x)$. We show that $\mathcal{M}_{nn}$ always mixes in polynomial time. As a corollary, we show that the transposition chain $\mathcal{M}_T$ that transposes any pair of objects (with appropriate transition probabilities to maintain the stationary distribution) is also rapidly mixing for this class of $\mathbf{P}$.

In the second class of $\mathbf{P}$ we consider, we are given parameters $1/2 \leq q_1, \ldots, q_{\log n} \leq 1$. Then for $x < y$, we define $p_{x,y} = q_i$ where i is the highest order bit in which the binary representations of x and y differ and as above $p_{y,x} = 1 - p_{x,y}$. In other words, if $n = 2^k$, then $p_{x,y} = q_1$ if $x \leq n/2$ and $y > n/2$. However, if $x \leq n/4$ and $n/4 < y \leq n/2$, then they have rate $p_{x,y} = q_2$. We can think of this setting as a crude representation of winning probabilities for tournaments: if two people are from competing sports leagues, and one league is stronger, then there is a fixed probability that the player from the stronger league wins the match. Likewise, within either of

the two leagues there are divisions A and B, and players from each of them have a fixed probability of winning if matched to anyone from the other division, and these divisions are then further subdivided recursively. We show the Markov chain $\mathcal{M}_T$ is rapidly mixing for any choice of the $\{q_i\}$. Moreover, we extend this analysis to a general class of $\mathbf{P}$ that have what we call *tree structure*. In this class, we have any binary tree with objects $1, \ldots, n$ on the leaves and $q_{v_1}, \ldots, q_{v_{n-1}}$ assigned to the internal nodes of the tree, with $1/2 \leq q_{v_i} \leq 1$, and for all $x < y$, we define $p_{x,y} = q_{x \vee y}$ where $x \vee y$ is the join of x and y in the tree. Both of these settings in this second class include permutations with constant bias as a special case, although we do not get the optimal bound achieved by Benjamini et al. [5]. Using the comparison theorem we have a corollary showing that $\mathcal{M}_{nn}$ is also rapidly mixing for this class if the q_i (or q_{v_i}) satisfy a "weak regularity" condition.

While this is encouraging evidence that the chain on biased permutations is always rapidly mixing, our final theorem presents a counterexample that disproves this conjecture in the most general setting for positively biased permutations. We do not know if the spectral gap is always bounded when $\mathbf{P}$ also satisfies the regularity condition and leave this as an open problem.

Techniques: All of our results are based on placing the problems in a new framework using insights from combinatorics and statistical physics. For each of our positive results, we use combinatorial bijections to reformulate the chain on permutations in terms of other combinatorial objects. These powerful bijections allow us to find Markov chains in the new setting that are straightforward to analyze, and then we use the comparison theorem to derive a bound on the mixing time of $\mathcal{M}_{nn}$. For the negative result, we use a combinatorial map from permutations to sets of lattice paths, and then use intuition from physics to determine when the chain $\mathcal{M}_{nn}$ can be slow.

For our first class of $\mathbf{P}$ values, where $p_{x,y}$ is determined by the smaller of x and

y, we use a bijection between permutations and their *inversion tables*. The ith entry of an inversion table is the number of integers larger than i that come before i in the permutation. We then introduce a new Markov chain $\mathcal{M}_{inv}$ that acts on inversion tables by changing one number on the table at a time. We show that $\mathcal{M}_{inv}$ is always rapidly mixing, and then we use a comparison argument to infer that $\mathcal{M}_{nn}$ is also. This gives a remarkably simple proof for $\mathbf{P}$ values in the first class, including a new proof of fast mixing in the constant bias case that is much simpler than all previous arguments. Also, in the second class, when $\mathbf{P}$ has tree structure, we give an alternative way to represent the permutations as collections of independent ASEP, in the spirit of the Benjamini et al. reduction. We represent the permutations as binary trees with each non-leaf node corresponding to sequences of zeros and ones, and we identify a Markov chain $\mathcal{M}_{tree}$ whose behavior within each non-leaf node is precisely an ASEP. The Benjamini et al. [5] result gives optimal bounds on the mixing time of the ASEP when the number of ones is nearly equal to the number of zeroes, and an improved analysis of the unbalanced case was given by Greenberg et al. [33]. We use the latter to improve our bound on the mixing time of $\mathcal{M}_{tree}$. We then compare the Markov chain $\mathcal{M}_{tree}$ with $\mathcal{M}_{nn}$ to show that $\mathcal{M}_{nn}$ is also rapidly mixing as long as $\mathbf{P}$ is weakly regular. Interestingly, this condition is very similar to the regularity condition described by Fill in [28].

For our negative result, the choice of $\mathbf{P}$ was motivated by a related question arising in the context of biased staircase walks [33]. In that context, we are sampling ASEP configurations with n zeros and n ones, which map bijectively onto walks on the Cartesian lattice from $(0, n)$ to $(n, 0)$ that always go to the right or down. The probability of each configuration w is proportional to $\Pi_{xy<w}\lambda_{xy}$, where $1/2 \leq \lambda_{xy}$ is assigned to the square at (x, y) and $xy < w$ whenever the square at (x, y) lies underneath the walk. We find in this context that there is a setting of the $\{\lambda_{xy}\}$ where each is close to $1/2$ or close to 1 for which there is a "disordered state" where the paths

concentrate near the diagonal from $(0, n)$ to $(n, 0)$ and an "ordered state" where paths come close to (n, n) and it is difficult to move between them (the disordered state has many configurations, all with small stationary probability, while the ordered state has many fewer configurations, all with high probability). This allows us to show that the chain is slowly mixing for staircase walks, and we can extend this result to determine values of $\mathbf{P}$ to show $\mathcal{M}_{nn}$ can be slowly mixing on permutations.

4.2 Model and techniques

We begin by formalizing our model. Let Ω be the set of all permutations $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$ of n integers. We consider Markov chains on Ω whose transitions transpose two elements of the permutation. We are also given a set $\mathbf{P}$, consisting of $p_{i,j} \in [0, 1]$ for each $1 \leq i \neq j \leq n$, where for any $i < j$, $p_{i,j} \geq 1/2$ and $p_{j,i} = 1 - p_{i,j}$. The Markov chain $\mathcal{M}_{nn}$ will sample from Ω using $\mathbf{P}$.

The Nearest Neighbor Markov chain $\mathcal{M}_{nn}$

`Starting at any permutation` σ_0`, iterate the following:`

- `At time` t`, select an index` $i \in [n-1]$ `uniformly at random (u.a.r).`
  - `Swap the elements` $\sigma_t(i), \sigma_t(i+1)$ `with probability` $p_{\sigma_t(i+1),\sigma_t(i)}$ `to obtain` σ_{t+1}`.`
  - `Do nothing with probability` $p_{\sigma_t(i),\sigma_t(i+1)}$ `so that` $\sigma_{t+1} = \sigma_t$`.`

The Markov chain $\mathcal{M}_{nn}$ connects the state space, since every permutation σ can move to the ordered permutation $(1, 2, \ldots, n)$ (and back) using the bubble sort algorithm. Since $\mathcal{M}_{nn}$ is also aperiodic, this implies that $\mathcal{M}_{nn}$ is ergodic. It is easy to see that for $\mathcal{M}_{nn}$, the distribution $\pi(\sigma) = \prod_{(i<j)} p_{\sigma(i),\sigma(j)}/Z$, where Z is the normalizing constant $\sum_{\sigma\in\Omega} \prod_{(i<j)} p_{\sigma(i),\sigma(j)}$, satisfies the detailed balance condition (Definition 2.1.1), and is thus the stationary distribution.

The Markov chain $\mathcal{M}_T$ can make any transposition at each step, while maintaining the stationary distribution π. The transition probabilities of $\mathcal{M}_T$ can be quite complicated, since swapping two distant elements in the permutation consists of many transitions of $\mathcal{M}_{nn}$, each with different probabilities. In the following sections, we will introduce two other Markov chains whose transitions are a subset of those of $\mathcal{M}_T$ for which we can describe the transition probabilities succinctly.

4.2.1 Analyzing a product of Markov chains

For each of our positive results, we show that the Markov chain in question can be decomposed into M independent Markov chains. Since each Markov chain $\mathcal{M}_i$ operates independently, the overall mixing time will be roughly M times the mixing time of each piece, slowed down by the inverse probability of selecting that process. This type of decomposition has been done before (e.g., see [7]) but in the more restricted case where we choose each component in the product with equal probability. We include the proof of the generalization for completeness.

Theorem 4.2.1. *Suppose the Markov chain $\mathcal{M}$ is a product of M independent Markov chains $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_M$, where $\mathcal{M}$ updates $\mathcal{M}_i$ with probability p_i, where $\sum_i p_i = 1$. If $\tau_i(\epsilon)$ is the mixing time for $\mathcal{M}_i$ and $\tau_i(\epsilon) \geq 4 \ln \epsilon$ for each i, then*

$$\tau(\epsilon) \leq \max_{i=1,2,\ldots,M} \frac{2}{p_i} \tau_i\left(\frac{\epsilon}{2M}\right).$$

Proof. Suppose the Markov chain $\mathcal{M}$ has transition matrix P, and each $\mathcal{M}_i$ has transition matrix P_i and state space Ω_i. Let $B_i = p_i P_i + (1 - p_i)I$, where I is the identity matrix of the same size as P_i, be the transition matrix of $\mathcal{M}_i$, slowed down by the probability p_i of selecting $\mathcal{M}_i$. First we show that the total variation distance satisfies

$$1 + 2d_{tv}(P^t, \pi) \leq \prod_i (1 + 2d_{tv}(B_i^t, \pi_i)).$$

To show this, notice that for $x = (x_1, x_2, \ldots, x_M), y = (y_1, y_2, \ldots, y_M) \in \Omega$, $P^t(x, y) = \prod_i B_i^t(x_i, y_i)$. Let $\epsilon_i(x_i, y_i) = B_i^t(x_i, y_i) - \pi_i(y_i)$ and for any $x_i \in \Omega_i$,

$$\epsilon_i(x_i) = \sum_{y_i \in \Omega_i} |\epsilon_i(x_i, y_i)| \leq 2 d_{tv}(B_i^t, \pi_i).$$

Then,

$$\begin{aligned}
d_{tv}(P^t, \pi) &= \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)| \\
&= \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} \left| \prod_i B_i^t(x_i, y_i) - \prod_i \pi_i(y_i) \right| \\
&= \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} \left| \prod_i \left(B_i^t(x_i, y_i) - \pi_i(y_i) + \pi_i(y_i) \right) - \prod_i \pi_i(y_i) \right| \\
&= \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} \left| \prod_i \left(\epsilon_i(x_i, y_i) + \pi_i(y_i) \right) - \prod_i \pi_i(y_i) \right| \\
&= \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} \left| \sum_{S \subseteq [M], S \neq \emptyset} \prod_{i \in S} \epsilon_i(x_i, y_i) \prod_{i \notin S} \pi_i(y_i) \right| \\
&\leq \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} \sum_{S \subseteq [M], S \neq \emptyset} \prod_{i \in S} |\epsilon_i(x_i, y_i)| \prod_{i \notin S} |\pi_i(y_i)| \\
&= \max_{x \in \Omega} \frac{1}{2} \sum_{S \subseteq [M], S \neq \emptyset} \prod_{i \in S} \sum_{y_i \in \Omega_i} |\epsilon_i(x_i, y_i)| \prod_{i \notin S} \sum_{y_i \in \Omega_i} |\pi_i(y_i)| \\
&= \max_{x \in \Omega} \frac{1}{2} \sum_{S \subseteq [M], S \neq \emptyset} \prod_{i \in S} \epsilon_i(x_i) \prod_{i \notin S} 1 \\
&= \max_{x \in \Omega} \frac{1}{2} \prod_i (1 + \epsilon_i(x_i)) - 1/2 \quad \leq \quad \frac{1}{2} \prod_i (1 + 2 d_{tv}(B_i^t, \pi_i)) - 1/2,
\end{aligned}$$

as desired. Thus in order to get $d_{tv}(P^t, \pi) \leq \epsilon$, it suffices to show $d_{tv}(B_i^t, \pi_i) \leq \epsilon/(2M)$ for each i, because then

$$\begin{aligned}
1 + 2 d_{tv}(P^t, \pi) \leq \prod_i (1 + 2 d_{tv}(B_i^t, \pi_i)) \ &\leq \ \prod_i (1 + 2\epsilon/(2M)) \\
&\leq \quad e^{\epsilon} \quad \leq \ 1 + 2\epsilon.
\end{aligned}$$

Hence it suffices to show $d_{tv}(B_i^t, \pi_i) \leq \epsilon/(2M)$ for each i.

Since $B_i^t = (p_i P_i + (1-p_i)I)^t = \sum_{j=0}^t \binom{t}{j} p_i^j (1-p_i)^{t-j} P_i^j I$, we have

$$\begin{aligned}
d_{tv}(B_i^t, \pi_i) &= \max_{x_i \in \Omega_i} \frac{1}{2} \sum_{y_i \in \Omega_i} \left| B_i^t(x_i, y_i) - \pi_i(y_i) \right| \\
&= \max_{x_i \in \Omega_i} \frac{1}{2} \sum_{y_i \in \Omega_i} \left| \sum_{j=0}^t \binom{t}{j} p_i^j (1-p_i)^{t-j} P_i^j(x_i, y_i) - \pi_i(y_i) \right| \\
&= \max_{x_i \in \Omega_i} \frac{1}{2} \sum_{y_i \in \Omega_i} \left| \sum_{j=0}^t \binom{t}{j} p_i^j (1-p_i)^{t-j} P_i^j(x_i, y_i) - \sum_{j=0}^t \binom{t}{j} p_i^j (1-p_i)^{t-j} \pi_i(y_i) \right| \\
&\leq \max_{x_i \in \Omega_i} \frac{1}{2} \sum_{y_i \in \Omega_i} \sum_{j=0}^t \binom{t}{j} p_i^j (1-p_i)^{t-j} \left| P_i^j(x_i, y_i) - \pi_i(y_i) \right| \\
&= \sum_{j=0}^t \binom{t}{j} p_i^j (1-p_i)^{t-j} \max_{x_i \in \Omega_i} \frac{1}{2} \sum_{y_i \in \Omega_i} \left| P_i^j(x_i, y_i) - \pi_i(y_i) \right| \\
&= \sum_{j=0}^t \binom{t}{j} p_i^j (1-p_i)^{t-j} d_{tv}(P_i^j, \pi_i).
\end{aligned}$$

Let $t_i = \tau_i(\epsilon/(4M))$. Now, for $j \geq t_i = \tau_i(\epsilon/(4M))$, we have that $d_{tv}(P_i^j, \pi_i) < \epsilon/(4M)$. For all j, we have $d_{tv}(P_i^j, \pi_i) \leq 2$, so if X is a binomial random variable with parameters t and p_i, we have

$$\begin{aligned}
d_{tv}(B_i^t, \pi_i) &\leq \sum_{j=0}^t \binom{t}{j} p_i^j (1-p_i)^{t-j} d_{tv}(P_i^j, \pi_i) \\
&= \sum_{j=0}^{t_i - 1} \binom{t}{j} p_i^j (1-p_i)^{t-j} d_{tv}(P_i^j, \pi_i) + \sum_{j=t_i}^t \binom{t}{j} p_i^j (1-p_i)^{t-j} d_{tv}(P_i^j, \pi_i) \\
&< 2 \sum_{j=0}^{t_i - 1} \binom{t}{j} p_i^j (1-p_i)^{t-j} + \sum_{j=t_i}^t \binom{t}{j} p_i^j (1-p_i)^{t-j} \epsilon/(2M) \\
&= 2P(X < t_i) + \epsilon/(2M).
\end{aligned}$$

By Chernoff bounds, $P(X < (1-\delta)tp_i) \leq e^{-tp_i\delta^2/2}$. Setting $\delta = 1 - t_i/(tp_i)$, then for all $t > 2t_i/p_i$, $\delta^2 \geq 1/4$ and we have

$$P(X < t_i) \leq e^{-tp_i\delta^2/2} \leq e^{-tp_i/8} \leq \epsilon/(8M),$$

as long as $t \geq 8\ln(\epsilon/(8M))/p_i$. Therefore for $t \geq \max\{8\ln(\epsilon/(8M))/p_i, 2t_i/p_i\}$,

$$d_{tv}(B_i^t, \pi_i) = 2P(X < t_i) + \epsilon/(4M) \ \leq \ 2\epsilon/(8M) + \epsilon/(4M) \ = \ \epsilon/(2M).$$

Hence by time t the total variation distance satisfies $d_{tv}(P^t, \pi) \leq \epsilon$. □

4.3 First class: Sampling inversion tables

In this section we assume we are given $1/2 \leq r_1, r_2, \ldots, r_{n-1} \leq 1$ and a set $\mathbf{P}$ satisfying $p_{i,j} = r_i$, if $i < j$ and $p_{i,j} = 1 - p_{j,i}$ if $j < i$. Note that this analysis extends to the case that $p_{i,j} = r_j$ for all $i < j$, which is the setting for self-organizing lists. We show that a new Markov chain $\mathcal{M}_{inv}$ is rapidly mixing under these conditions, which will imply that $\mathcal{M}_{nn}$ and $\mathcal{M}_T$ are as well, as we show in Section 4.5. The Markov chain $\mathcal{M}_{inv}$ acts on the *inversion table* of the permutation, which has an entry for each $i \in [n]$ counting the number of inversions involving i; that is, the number of values $j > i$ where j comes before i in the permutation (see Figure 10). It is easy to see that the ith element of the inversion table is an integer between 0 and $n - i$. In fact, the function I is a bijection between the set of permutations and the set $\mathcal{I}$ of all possible inversion tables (all sequences $X = (x_1, x_2, \ldots, x_n)$ where $0 \leq x_i \leq n - i$ for all $i \in [n]$). To see this, we will construct a permutation from any inversion table $X \in I$. Place the element 1 in the $(x_1 + 1)$st position of the permutation. Next, there are $n - 1$ slots remaining. Among these, place the element 2 in the $(x_2 + 1)$st position remaining (ignoring the slot already filled by 1). Continuing, after placing $i - 1$ elements into the permutation, there are $n - i + 1$ slots remaining, and we place the element i into the $(x_i + 1)$st position among the remaining slots. This proves that I is a bijection from S_n to $\mathcal{I}$.

σ	=	8	1	5	3	7	4	6	2
$I(\sigma)$	=	1	7	2	3	1	2	1	0

Figure 10: The inversion table for a permutation.

Given this bijection, a natural algorithm for sampling permutations is to perform the following local Markov chain on inversion tables: select a position $i \in [n]$ and attempt to either add one or subtract one from x_i, according to the appropriate

probabilities. In terms of permutations, this amounts to adding or removing an inversion involving i without affecting the number of inversions involving any other integer, and is achieved by swapping the element i with an element $j > i$ such that every element in between is smaller than both i and j. If i moves ahead of j, this move happens with probability $p_{i,j}$ because for each k that i and j are swapped past, $p_{k,i} = r_k = p_{k,j}$, so the net effect on the distribution is neutral, and the detailed balance condition ensures that π is the correct stationary distribution. Formally, the Markov chain is defined as follows.

The Inversion Markov chain $\mathcal{M}_{inv}$

Starting at any permutation σ_0, iterate the following:

- Select an element $i \in [n]$ with probability $(n-i)/\binom{n}{2}$ and a bit $b \in \{-1, +1\}$.
  - If $b = +1$, let j be the first element after element i in σ_t such that $j > i$. With prob. $p_{j,i}/2 = (1 - r_i)/2$, obtain σ_{t+1} from σ_t by swapping i and j.
  - If $b = -1$, let j be the last element before element i in σ_t such that $j > i$. With prob. $p_{i,j}/2 = r_i/2$, obtain σ_{t+1} from σ_t by swapping i and j.
- With prob. $1/2$, $\sigma_{t+1} = \sigma_t$.

This Markov chain contains the moves of $\mathcal{M}_{nn}$ (and therefore also connects the state space). Although elements can jump across several elements, it is still fairly local compared with the general transposition chain $\mathcal{M}_T$ which has $\binom{n}{2}$ choices at every step, since $\mathcal{M}_{inv}$ has at most $2n$.

The Markov chain $\mathcal{M}_{inv}$ is essentially a product of n independent one-dimensional processes. The ith process is just a random walk bounded between 0 and $n-i$, which

moves up with probability $1 - r_i$ and down with probability r_i; hence its mixing time is $O(n^2)$, unless r_i is bounded away from $1/2$, in which case its mixing time is $O(n)$. However, each process is slowed down by a factor of n since we only update one process at each step.

The key here is to notice that the one-dimensional processes are still independent in the biased case, as long as the bias depends only on the smaller of the two integers. This is because when $\mathcal{M}_{inv}$ swaps i and j in the permutation, where $i < j$, it affects the inversion table only in the ith coordinate. Therefore, the inversion table is a product of n independent one-dimensional processes, each of which will converge in time $O(n)$, and since again it is slowed down by a factor of n, we will get a bound of $O(n^2)$, as long as each r_i is a constant.

Theorem 4.3.1. *Let $1/2 \leq r_1, r_2, \ldots, r_{n-1} < 1$ be constants, and let $r_{max} = \max_i r_i$. Assume that $p_{i,j} = r_{\min\{i,j\}}$.*

1. *If each $r_i > 1/2$ then the mixing time of $\mathcal{M}_{inv}$ on biased permutations with these $p_{i,j}$ values is $O(n^2 \ln(n/\epsilon))$.*

2. *Otherwise, the mixing time of $\mathcal{M}_{inv}$ is $O(n^3 \ln(n/\epsilon))$.*

To prove this theorem, we will use Theorem 4.2.1. First, we need to analyze the one-dimensional process $\mathcal{M}(r, k)$, bounded between 0 and k, which chooses to move up with probability $r \geq 1/2$ and down with probability $1 - r$ at each step, if possible. This simple random walk is well-studied; we include the proof for completeness.

Lemma 4.3.2. *Let $1/2 \leq r \leq 1$ be constant. Then the Markov chain $\mathcal{M}(r, k)$ has mixing time*

1. *$\tau(\epsilon) = O(k \ln \epsilon^{-1})$ if r is a constant bigger than $1/2$, and*

2. *$\tau(\epsilon) = O(k^2 \ln \epsilon^{-1})$ if $r = 1/2$.*

Proof. We use a variation on coupling. We use the trivial coupling, which chooses to move the same direction in each Markov chain. Notice that the Markov chain $\mathcal{M}(r,k)$ is monotone with respect to this coupling, in the sense that if X_t is below Y_t, then it will remain so until $X_{t'} = Y_{t'}$. Thus the time until the chains couple is bounded by the time it takes for a process Z_t, where $Z_0 = 0$, to reach height k. However, Z_t is just a biased random walk bounded between 0 and k. First, we notice that Z_t is non-decreasing in expectation; that is, for all $t > 0$, $E[Z_{t+1} - Z_t] \geq 0$:

$$E[Z_{t+1} - Z_t] \quad = \quad r - (1-r) = 2r - 1 \geq 0.$$

Consider the case that $r > 1/2$. Define $W(t) = k - Z(t) + (2r-1)t$. Examining the expected difference between $W(t)$ and $W(t+1)$, we see

$$E[W(t+1) - W(t)] = E[-Z(t+1) + 2r - 1 + Z(t)] = 0.$$

Also, since the differences $W(t+1) - W(t)$ are bounded, $\{W(t)\}$ is a martingale. The time $T = \min\{t : Z_t = 0\}$ is a stopping time for the process $W(t)$, so we may apply the Optional Stopping Theorem for martingales to deduce that

$$E[W(T)] = W(0) = k.$$

However, since

$$E[W(T)] = E[k - Z(T) + (2r-1)T] = (2r-1)E[T],$$

it follows that $E[T] = k/(2r-1)$. Recall from Theorem 2.2.2 that

$$\tau(\epsilon) = O(T \ln \epsilon^{-1}) = O(k/(2r-1) \ln \epsilon^{-1}) = O(k \ln \epsilon^{-1}).$$

Suppose now that $r = 1/2$. This case is similar, and follows from Lemma 6 of [49]. Notice $E[(Z(t+1) - Z(t))^2] = r + (1-r) = 1$. Therefore $E[T] \leq k(2k-k)/1 = k^2$. Hence $\tau(\epsilon) = O(k^2 \ln \epsilon^{-1})$.

□

Finally, we can use these bounds to prove Theorem 4.3.1.

Proof of Theorem 4.3.1. The ith process is chosen with probability $(n-i)/(2\binom{n}{2})$. Therefore, by Theorem 4.2.1, the mixing time of $\mathcal{M}_{inv}$ satisfies

$$\tau(\epsilon) \leq \frac{\binom{n}{2}}{n-i}(n-i)\ln(2n/\epsilon) = \binom{n}{2}\ln(2n/\epsilon) = O(n^2\ln(n/\epsilon))$$

when each r_i is bounded away from $1/2$. Otherwise,

$$\tau(\epsilon) \leq \frac{\binom{n}{2}}{n-i}(n-i)^2\ln(2n/\epsilon) = O(n^3\ln(n/\epsilon)).$$

□

4.4 Second class: The Markov chain $\mathcal{M}_{tree}(T)$

In this section, we turn to a second class of $\mathbf{P}$ that have what we call *tree structure*. Let T be a proper rooted binary tree with n leaf nodes, labeled $1, \ldots, n$ in sorted order. Each non-leaf node v of this tree is labeled with a value $\frac{1}{2} \leq q_v \leq 1$. For $i, j \in [n]$, let $i \vee j$ be the lowest common ancestor of the leaves labeled i and j. We say that $\mathbf{P}$ *has tree structure* T if for all $i < j$, $p_{i,j} = q_{i \vee_T j}$ and $p_{j,i} = 1 - p_{i,j}$. For example, Figure 11a shows a set $\mathbf{P}$ such that $p_{14} = .8$, $p_{49} = .9$, and $p_{58} = .7$.

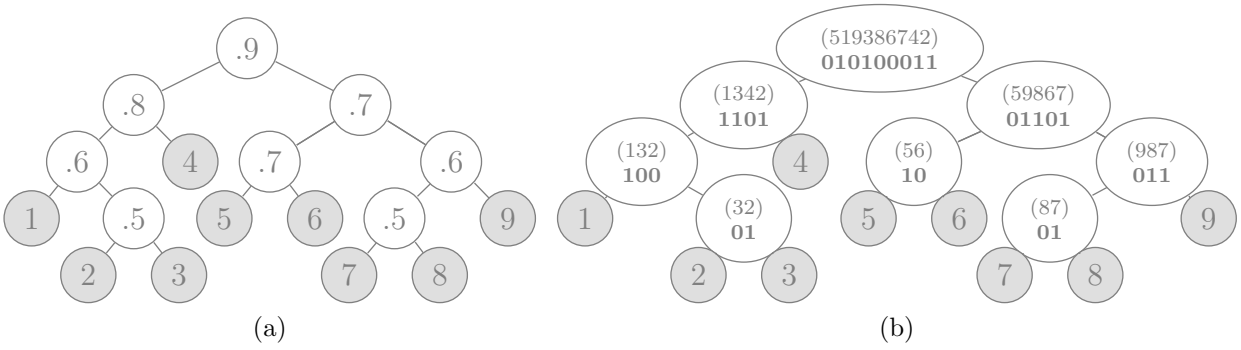


Figure 11: A set $\mathbf{P}$ with tree structure, and the corresponding tree encoding of the permutation 519386742.

When T is a complete binary tree and $q_{v_1} = q_{v_2}$ for each v_1 and v_2 on the same level of the tree, this is precisely the representation of the winning probabilities for

a tournament described in the introduction. We define the Markov chain $\mathcal{M}_{tree}(T)$ over permutations, given a set $\mathbf{P}$ with tree structure T.

The Markov chain $\mathcal{M}_{tree}(T)$

Starting at any permutation σ_0, iterate the following:

- Select distinct $a, b \in [n]$ u.a.r. Assume $a < b$.
- If every number between a and b in the permutation σ_t is not a descendant in T of $a \vee_T b$, obtain σ_{t+1} from σ_t by placing a, b in order with probability $p_{a,b}$, and out of order with probability $1-p_{a,b}$, leaving all elements between them fixed.
- Otherwise, $\sigma_{t+1} = \sigma_t$.

First, we show that this Markov chain samples from the same distribution as $\mathcal{M}_{nn}$. Swapping arbitrary non-adjacent elements a and b could potentially change the weight of the permutation dramatically. However, for any element c that is not a descendant in T of $a \vee_T b$, the relationship between a and c is the same as the relationship between b and c. Thus the tree structure ensures that swapping a and b only changes the weight by a multiplicative factor of $\lambda_{a,b} = p_{a,b}/p_{b,a}$.

Lemma 4.4.1. *The Markov chain $\mathcal{M}_{tree}(T)$ has the same stationary distribution as $\mathcal{M}_{nn}$.*

Proof. Let π be the stationary distribution of $\mathcal{M}_{nn}$, and let (σ_1, σ_2) be a transition in $\mathcal{M}_{tree}(T)$. It suffices to show that the detailed balance condition holds for this transition with the stationary distribution π. Recall that we may express $\pi(\sigma) = \prod_{i,j|i<_\sigma j} p_{i,j}/Z$ where $Z = \sum_{\sigma\in\Omega} \prod_{i,j|i<_\sigma j} p_{i,j}$. The transition (σ_1, σ_2) transposes some two elements $a <_{\sigma_1} b$, where every element between a and b in σ_i is not a descendant of $a \vee b$ in T. Let $x_1, \ldots, x_k$ be those elements. Thus, the path from a or b to x_i in T

must pass through $a \vee b$ and go to another part of the tree. For every such element x_i, $a \vee x_i = (a \vee b) \vee x_i = b \vee x_i$. From the observation, we see from the tree structure that $p_{ax_i} = p_{bx_i}$ for every x_i between a and b. In particular, x_i is either greater than both a and b or less than both a and b, since all integers c such that $a < c < b$ are necessarily descendants of $a \vee b$. Therefore,

$$\frac{\pi(\sigma_1)}{\pi(\sigma_2)} = \frac{p_{ab} \prod_i p_{ax_i}}{p_{ba} \prod_i p_{bx_i}} = \frac{p_{ab}}{p_{ba}}.$$

This is exactly the ratio of the transition probabilities in $\mathcal{M}_{tree}(T)$, thus $\mathcal{M}_{tree}(T)$ also has stationary distribution π. □

The key to the proof that $\mathcal{M}_{tree}(T)$ is rapidly mixing is again to decompose the chain into $n-1$ independent Markov chains, $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_{n-1}$, one for each non-leaf node of the tree T. To this end, we introduce an alternate representation of a permutation as a set of binary strings arranged like the tree T. For each non-leaf node v in the tree T, let $L(v)$ be its left descendants, and $R(v)$ be its right descendants. We now do the following: Given the permutation σ, list each descendant x of v in the order we encounter it in σ; these are parenthesized in Figure 11*b*. Then for each listed element x, write a 1 if $x \in L(v)$ and a 0 if $x \in R(v)$. This is the final binary encoding in Figure 11*b*. We see that any σ will lead to an assignment of binary strings at each non-leaf node v with $L(v)$ ones and $R(v)$ zeroes. Next we verify that this is a bijection between the set of permutations and the set of assignments of such binary strings to the tree T. Given any such assignment of binary strings, we can recursively reconstruct the permutation σ as follows. For each leaf node i, let its string be the string "i". For any node n with binary string b, determine the strings of its two children. Call these s_1, s_0. Interleave the elements of s_1 with s_0, choosing an element of s_1 for each 1 in b, and an element of s_0 for each 0. This yields a permutation σ.

With this bijection, we first analyze $\mathcal{M}_{tree}(T)$'s behavior over tree representations and later extend this analysis to permutations. The Markov chain $\mathcal{M}_{tree}(T)$, when

proposing a swap of the elements a and b, will only attempt to swap them if a, b correspond to some *adjacent* 0 and 1 in the string associated with $a \vee b$. Swapping a and b does not affect any other string, so each non-leaf node v represents an independent exclusion process with $L(v)$ ones and $R(v)$ zeroes. These exclusion processes are precisely the staircase walks analyzed in Chapter 3. We will use the following bounds on the mixing times of the symmetric and asymmetric exclusion processes; the first is from Bubley and Dyer [11] and Wilson [67], and the second is from the analysis in Chapter 3.

Theorem 4.4.2. *Let $\mathcal{M}$ be the exclusion process with parameter p on k_1 ones and k_2 zeroes, where $k = k_1 + k_2$. Then*

1. *if $p = 1/2$, $\tau(\epsilon) = O(k^3 \log(k_1 k_2/\epsilon))$.*
2. *if $p > 1/2$, then $\tau(\epsilon) = O(k(\min\{k_1, k_2\} + \log k)\log(\epsilon^{-1})) = O(k^2 \log(\epsilon^{-1}))$.*

The bounds in Theorem 4.4.2 refer to the exclusion process which selects a position at random and swaps the two elements in that position with the appropriate probability. However, our process selects arbitrary pairs (i, j) consisting of a single one and a single zero. Since we only swap (i, j) if they are neighboring, this may slow down the chain by a factor of at most k.

Since each exclusion process operates independently, the overall mixing time will be roughly n times the mixing time of each piece, slowed down by the inverse probability of selecting that process. Next, we will use Theorems 4.2.1 and 4.4.2 to prove that $\mathcal{M}_{tree}(T)$ is rapidly mixing.

Theorem 4.4.3. *If $\mathbf{P}$ has tree structure T, then the mixing time of $\mathcal{M}_{tree}(T)$ under $\mathbf{P}$ satisfies*

$$\tau_{tree}(\epsilon) = O(n^5 \log(n/\epsilon)).$$

If $\mathbf{P}$ is such that each $q_i > 1/2$ is a constant, then $\tau_{tree}(\epsilon) = O(n^3 \log n \log(n/\epsilon))$.

Proof. In order to apply Theorem 4.2.1 to the Markov chain $\mathcal{M}_{tree}(T)$, we note that for a node with k_1 ones and k_2 zeroes ($k = k_1 + k_2$), the probability of selecting that node is $k_1 k_2 / \binom{n}{2}$. Since $M = n - 1$, Theorem 4.2.1 implies

$$\tau(\epsilon) \leq \frac{n(n-1)}{k_1 k_2} k^4 \ln(2nk_1k_2/\epsilon) = O(n^5 \log(n/\epsilon)).$$

Of course, if all of the chains have probabilities that are bounded away from 1/2, then we can use the second bound from Theorem 4.4.2 to obtain

$$\begin{aligned}\tau(\epsilon) &\leq \frac{n(n-1)}{k_1 k_2} k^2 (\min\{k_1, k_2\} + \log k) \log(2n/\epsilon) \\ &\leq \frac{n(n-1)k^2}{\max\{k_1, k_2\}} \left(1 + \frac{\log k}{\min\{k_1, k_2\}}\right) \log(2n/\epsilon).\end{aligned}$$

There are two cases to consider. Let $0 < c < 1$. If $\min\{k_1, k_2\} \geq c \log k$ then

$$\tau(\epsilon) \leq \frac{n(n-1)k^2}{k/2}(1 + c) \log(2n/\epsilon)) = O(n^3 \log(n/\epsilon)).$$

Otherwise, $\max\{k_1, k_2\} > k - c \log k$, so since $k \leq n$,

$$\tau(\epsilon) \leq \frac{n(n-1)k^2}{k - c \log k}(1+\log k) \log(2n/\epsilon)) = \frac{n(n-1)k}{1 - \frac{c \log k}{k}}(1+\log k) \log(2n/\epsilon)) = O(n^3 \log n \log(n/\epsilon)).$$

□

4.5 Bounding the mixing time of $\mathcal{M}_{nn}$ for both classes

Our goal now is to use the comparison method to obtain bounds on the mixing time of $\mathcal{M}_{nn}$ in the settings of Sections 4.3 and 4.4 from the bounds on the mixing times of $\mathcal{M}_{inv}$ and $\mathcal{M}_{tree}(T)$. When comparing the mixing times of $\mathcal{M}_{tree}(T)$ and $\mathcal{M}_{nn}$, for example, the goal is to show that a move $e = (\sigma, \beta)$ of $\mathcal{M}_{tree}(T)$, which is allowed to transpose i and j that are not necessarily nearest neighbors, can be simulated with a sequence of moves of $\mathcal{M}_{nn}$. Moreover, we must ensure that our path does not go through transitions that are much smaller in weight than $\min\{\pi(\sigma), \pi(\beta)\}$. This type of argument is straightforward for the moves of $\mathcal{M}_{inv}$, and gives some intuition for

the more involved argument to compare $\mathcal{M}_{tree}(T)$ with $\mathcal{M}_{nn}$, which will follow in Section 4.5.2.

In the next two sections, we assume that each $p_{i,j}$ is a constant less than 1; this is to ensure a good comparison between the spectral gap and the mixing time. If this condition is not satisfied, then the proofs still go through and will give a bound on the spectral gap, but will not provide a good bound on the mixing time.

4.5.1 Comparing $\mathcal{M}_{inv}$ with $\mathcal{M}_{nn}$

First, we consider the setting of Section 4.3, where $p_{i,j}$ depends on $\min\{i, j\}$.

Theorem 4.5.1. *Let $1/2 \leq r_1, r_2, \ldots, r_{n-1} < 1$ be constants. Assume $\mathbf{P}$ is defined by $p_{i,j} = r_i$ for $i < j$. Then the mixing time of $\mathcal{M}_{nn}$ on biased permutations under $\mathbf{P}$ is $O(n^8 \log(n/\epsilon))$.*

Here we are using the bound from Theorem 4.3.1 part 2, and if each $p_{i,j}$ is bounded away from 1/2 then we would get a better bound of $O(n^7 \log(n/\epsilon))$ using Theorem 4.3.1 part 1. Recall that for any $a, b \in [n]$, we defined $\lambda_{a,b} = p_{a,b}/p_{b,a}$.

Proof. In order to apply Theorem 2.2.6, we need to define, for any transition $e = (\sigma, \beta)$ of the Markov chain $\mathcal{M}_{inv}$, a sequence of transitions of $\mathcal{M}_{nn}$. Let e be a transition of $\mathcal{M}_{inv}$ which performs a transposition on elements $\sigma(i)$ and $\sigma(j)$, where $i < j$. Recall $\mathcal{M}_{inv}$ can only swap $\sigma(i)$ and $\sigma(j)$ if all the elements between them are smaller than both $\sigma(i)$ and $\sigma(j)$. To obtain a sufficient bound on the congestion along each edge, we ensure that in each step of the path, we do not decrease the weight of the configuration. This is easy to do; in the first stage, move $\sigma(i)$ to the right, one step at a time, until it swaps with $\sigma(j)$. This removes an inversion of the type $(\sigma(i), \sigma(k))$ for every $i < k < j$, so clearly we have not decreased the weight of the configuration at any step. Next, move $\sigma(j)$ to the left, one step at a time, until it reaches position i. This completes the move e, and at each step, we are adding back an inversion of the type $(\sigma(j), \sigma(k))$ for some $i < k < j$. Since $\sigma(k) = \min\{\sigma(j), \sigma(k)\} = \min\{\sigma(i), \sigma(k)\}$,

we have $p_{\sigma(k),\sigma(i)} = p_{\sigma(k),\sigma(j)}$ for every $i < k < j$, so in this stage we restore all the inversions destroyed in the first stage, for a net change of $\lambda_{\sigma(j),\sigma(i)}$.

Given a transition (υ, ω) of $\mathcal{M}_{nn}$ we must upper bound the number of canonical paths $\gamma_{\sigma\beta}$ that use this edge, which we do by bounding the amount of information needed in addition to (υ, ω) to determine σ and β uniquely. For moves in the first stage, all we need to remember is $\sigma(j)$, because we know $\sigma(i)$ (it is the element moving forward). We also need to remember i (that is, the original location of $\sigma(i)$). Given this information along with υ and ω we can uniquely recover (σ, β). Thus there are at most n^2 paths which use any edge (υ, ω). Also, notice that the maximum length of any path is $2n$.

Next we bound the quantity A which is needed to apply Theorem 2.2.6. Recall that we have guaranteed that $\pi(\sigma) \leq \max\{\pi(\upsilon), \pi(\omega)\}$. Assume first that $\pi(\sigma) \leq \pi(\upsilon)$. Then

$$\begin{aligned} A &= \max_{(\upsilon,\omega)\in E(P)} \left\{ \frac{1}{\pi(\upsilon)P(\upsilon,\omega)} \sum_{\Gamma(\upsilon,\omega)} |\gamma_{\sigma\beta}|\pi(\sigma)P'(\sigma,\beta) \right\} \\ &\leq \max_{(\upsilon,\omega)\in E(P)} \sum_{\Gamma(\upsilon,\omega)} 2n \frac{P'(\sigma,\beta)}{P(\upsilon,\omega)} \;\leq\; \max_{(\upsilon,\omega)\in E(P)} \sum_{\Gamma(\upsilon,\omega)} 2n \frac{1/(2n)}{\frac{\lambda}{(1+\lambda)(n-1)}} \;=\; O(n^3). \end{aligned}$$

If, on the other hand, $\pi(\sigma) \leq \pi(\omega)$, then we use detailed balance to obtain:

$$\begin{aligned} A &= \max_{(\upsilon,\omega)\in E(P)} \left\{ \frac{1}{\pi(\upsilon)P(\upsilon,\omega)} \sum_{\Gamma(\upsilon,\omega)} |\gamma_{\sigma\beta}|\pi(\sigma)P'(\sigma,\beta) \right\} \\ &= \max_{(\upsilon,\omega)\in E(P)} \left\{ \frac{1}{\pi(\omega)P(\omega,\upsilon)} \sum_{\Gamma(\upsilon,\omega)} |\gamma_{\sigma\beta}|\pi(\sigma)P'(\sigma,\beta) \right\} \\ &\leq \max_{(\upsilon,\omega)\in E(P)} \sum_{\Gamma(\upsilon,\omega)} 2n \frac{P'(\sigma,\beta)}{P(\omega,\upsilon)} \\ &\leq \max_{(\upsilon,\omega)\in E(P)} \sum_{\Gamma(\upsilon,\omega)} 2n \frac{1/(2n)}{\frac{\lambda}{(1+\lambda)(n-1)}} \;=\; O(n^3). \end{aligned}$$

In either case, we have $A = O(n^3)$. Let $\lambda = \min_{i<j} \lambda_{j,i}$. Then $\pi_* = \min_{\rho\in\Omega} \pi(\rho) \geq \lambda^{\binom{n}{2}}/n!$, so $\log(1/(\epsilon\pi_*)) = O(n^2 \log \epsilon^{-1})$, since each $p_{i,j}$ bounded away from 1 implies

λ is a positive constant. Applying Theorem 2.2.6 proves that $\tau_{nn}(\epsilon) = O(n^8 \log(n/\epsilon))$.

□

4.5.2 Comparing $\mathcal{M}_{tree}(T)$ with $\mathcal{M}_{nn}$

In this section we show that $\mathcal{M}_{nn}$ is rapidly mixing when $\mathbf{P}$ has tree structure and is *weakly regular*:

Definition 4.5.1. *The set* $\mathbf{P}$ *is* weakly regular *if properties 1 and* either *2 or 3 are satisfied.*

1. $p_{i,j} \geq 1/2$ *for all* $1 \leq i < j \leq n$, *and*
2. $p_{i,j+1} \geq p_{i,j}$ *for all* $1 \leq i < j \leq n-1$ *or*
3. $p_{i-1,j} \geq p_{i,j}$ *for all* $2 \leq i < j \leq n$.

We note that if $\mathbf{P}$ satisfies all three properties then it is *regular*, as defined by Fill [28].

The comparison proof in this setting is similar to that of Section 4.5.1, except that there may be elements between $\sigma(i)$ and $\sigma(j)$ that are larger than both and elements that are smaller than both. This poses a problem, because we may not be able to move $\sigma(i)$ past all the elements between them without greatly decreasing the weight. However, when $\mathbf{P}$ is weakly regular, we can introduce a trick to get around this problem. At a high level, we shift the elements between $\sigma(i)$ and $\sigma(j)$ that are smaller than $\sigma(i)$ and $\sigma(j)$ to the left in a special way, increasing the weight of the configuration in such a way that when we move $\sigma(i)$ to the right, the weight never goes below $\min\{\pi(\sigma), \pi(\beta)\}$. Specifically, we prove the following theorem.

Theorem 4.5.2. *If* $\mathbf{P}$ *has tree structure, is weakly regular and is such that* $p_{i,j}$ *is a constant less than 1 for all* i, j, *then the mixing time of* $\mathcal{M}_{nn}$ *satisfies* $\tau_{nn}(\epsilon) = O(n^9 \log(n/\epsilon))$.

Again, we are assuming the worst case bound on the mixing time of $\mathcal{M}_{tree}(T)$ given in Theorem 4.4.3, and if each $p_{i,j}$ is bounded away from 1/2 then we would get a better bound.

Proof. Throughout this proof we assume that $\mathbf{P}$ satisfies properties 1 and 2 of the weakly regular definition. If instead $\mathbf{P}$ satisfies property 3, then the proof is very similar. In order to apply Theorem 2.2.6 to relate the mixing time of $\mathcal{M}_{nn}$ to the mixing time of $\mathcal{M}_{tree}(T)$ we need to define for each transition of $\mathcal{M}_{tree}(T)$ a canonical path using transitions of $\mathcal{M}_{nn}$. Let $e = (\sigma, \beta)$ be a transition of $\mathcal{M}_{tree}(T)$ which performs a transposition of elements $\sigma(i)$ and $\sigma(j)$ where $i < j$. If there are no elements between $\sigma(i)$ and $\sigma(j)$ then e is already a transition of $\mathcal{M}_{nn}$ and we are done. Otherwise, σ contains the string $\sigma(i), \sigma(i+1), ...\sigma(j-1), \sigma(j)$ and β contains $\sigma(j), \sigma(i+1), ...\sigma(j-1), \sigma(i)$. From the definition of $\mathcal{M}_{tree}(T)$ we know that for each $\sigma(k)$, $k \in [i+1, j-1]$, either $\sigma(k) > \sigma(i), \sigma(j)$ or $\sigma(k) < \sigma(i), \sigma(j)$. Define $S = \{\sigma(k) : \sigma_k < \sigma(i), \sigma(j)\}$ and $B = \{\sigma(k) : \sigma_k > \sigma(i), \sigma(j)\}$. To obtain a good bound on the congestion along each edge we must ensure that the weight of the configurations on the path are not smaller than the weight of σ. To this end, we define three stages in our path from σ to β. In the first, we shift the elements of S to the left, removing an inversion with each element of B. In the second stage we move $\sigma(i)$ next to $\sigma(j)$ and in the third stage we move $\sigma(j)$ to $\sigma(i)$'s original location. Finally, we shift the elements of S to the right to return them to their original locations. See Figure 12.

Stage 1: In this stage, for each $b \in B$, we remove an inversion involving b by shifting an element of S to the left past b. More precisely, if $\sigma(j-1) \in B$, shift $\sigma(j)$ to the left until an element from S is immediately to the left of $\sigma(j)$. Next, starting at the right-most element in S and moving left, for each $\sigma(k) \in S$ such that $\sigma(k-1) \in B$, move $\sigma(k)$ to the left one swap at a time until $\sigma(k)$ has an element from S or $\sigma(i)$ on its immediate left (see Figure 13a). Notice that for each element $b \in B$ we have

5	8	9	$\underline{2}$	10	$\underline{3}$	$\underline{4}$	$\underline{1}$	**7**
5	$\underline{2}$	8	9	$\underline{3}$	10	$\underline{4}$	$\underline{1}$	**7**
$\underline{2}$	8	9	$\underline{3}$	10	$\underline{4}$	$\underline{1}$	**5**	**7**
$\underline{2}$	8	9	$\underline{3}$	10	$\underline{4}$	$\underline{1}$	**7**	**5**
7	$\underline{2}$	8	9	$\underline{3}$	10	$\underline{4}$	$\underline{1}$	**5**
7	8	9	$\underline{2}$	10	$\underline{3}$	$\underline{4}$	$\underline{1}$	**5**

Figure 12: The canonical path for transposing 5 and 7. Notice that the elements in S are underlined.

removed exactly one $(b, \sigma(k))$ inversion where $\sigma(k) \in S \cup \sigma(j)$.

Stage 2: Next perform a series of nearest neighbor swaps to move $\sigma(i)$ to the right until it is in position j (the original position occupied by $\sigma(j)$ in σ, see Figure 13b). While we have created a $(b, \sigma(i))$ inversion for each element $b \in B$, we claim that the weight has not decreased from the original weight by more than a factor of $\lambda_{\sigma(j),\sigma(i)}$. This is because in Stage 1, for each element $b \in B$, we removed a (b, s) inversion for some $s \in S \cup \sigma(j)$. Assume first that $s \in S$. Then since $b > \sigma(i) > s$, it follows that $p_{b,\sigma(i)} \geq p_{b,s}$ for all $s \in S$ since the $\mathbf{P}$ are weakly regular; thus, for each b we introduce a multiplicative factor of $\lambda_{b,\sigma(i)}/\lambda_{b,s} \geq 1$. On the other hand, if $s = \sigma(j)$ then recall $p_{b,\sigma(j)} = p_{b,\sigma(i)}$ because b is not a descendant of $\sigma(i) \vee \sigma(j)$ in the tree T. Hence the current configuration has weight at least $\lambda_{\sigma(j),\sigma(i)}\pi(\sigma)$. Since $\lambda_{\sigma(j),\sigma(i)}$ is also the ratio of $\pi(\sigma)$ and $\pi(\beta)$, it follows that the weight at every step of Stage 2 does not go below $\min\{\pi(\sigma), \pi(\beta)\}$. For each $\sigma(k) \in S$ we have also removed a $(\sigma(k), \sigma(j))$ inversion, which can only increase the weight of the configuration.

5	8	9	$\underline{2}$	10	$\underline{3}$	$\underline{4}$	$\underline{1}$	**7**
5	8	9	$\underline{2}$	$\underline{3}$	10	$\underline{4}$	$\underline{1}$	**7**
5	8	$\underline{2}$	9	$\underline{3}$	10	$\underline{4}$	$\underline{1}$	**7**
5	$\underline{2}$	8	9	$\underline{3}$	10	$\underline{4}$	$\underline{1}$	**7**

(a)

5	$\underline{2}$	8	9	$\underline{3}$	10	$\underline{4}$	$\underline{1}$	**7**
$\underline{2}$	**5**	8	9	$\underline{3}$	10	$\underline{4}$	$\underline{1}$	**7**
$\underline{2}$	8	**5**	9	$\underline{3}$	10	$\underline{4}$	$\underline{1}$	**7**
				$\vdots$				
$\underline{2}$	8	9	$\underline{3}$	10	$\underline{4}$	$\underline{1}$	**5**	**7**
$\underline{2}$	8	9	$\underline{3}$	10	$\underline{4}$	$\underline{1}$	**7**	**5**

(b)

Figure 13: Stages 1 and 2 of the canonical path for transposing 5 and 7.

Stage 3: Perform a series of nearest neighbor swaps to move $\sigma(j)$ to the left until it is in the same position $\sigma(i)$ was originally. While we created an $(\sigma(k), \sigma(j))$ inversion for each $\sigma(k) \in S$, these inversions have the same weight as the $(\sigma(i), \sigma(k))$ inversion we removed in Stage 2. In addition we have removed an $(\sigma(l), \sigma(j))$ inversion for each $\sigma(l) \in B$.

Stage 4: Finally we want to return the elements in S and B to their original position. Starting with the left-most element in S that was moved in Stage 1, perform the nearest neighbor swaps to the right necessary to return it to its original position.

It's clear from the definition of the stages that along any path the weight of a configuration never decreases below the weight of $\min(\pi(\sigma), \pi(\beta))$. Given a transition (υ, ω) of $\mathcal{M}_{nn}$ we must upper bound the number of canonical paths $\gamma_{\sigma\beta}$ that use this edge. Thus, we analyze the amount of information needed in addition to (z, w) to determine σ and β uniquely. First we record whether (σ, β) is already a nearest-neighbor transition or which stage we are in. Next for any of the 4 stages we record the original location of $\sigma(i)$ and $\sigma(j)$. Given this information, along with υ and ω, we can uniquely recover (σ, β). Hence, there are at most $4n^2$ paths through any edge (υ, ω). Also, note that the maximum length of any path is $4n$.

Next we bound the quantity A which is needed to apply Theorem 2.2.6. Recall that for each transition (υ, ω) of the path $\gamma_{\sigma,\beta}$, we have guaranteed that $\pi(\upsilon) \geq \min\{\pi(\sigma), \pi(\beta)\}$. Assume first that $\pi(\upsilon) \geq \pi(\sigma)$. Then

$$\begin{aligned} A &= \max_{(\upsilon,\omega)\in E(P)} \left\{ \frac{1}{\pi(\upsilon)P(\upsilon,\omega)} \sum_{\Gamma(\upsilon,\omega)} |\gamma_{\sigma\beta}|\pi(\sigma)P'(\sigma,\beta) \right\} \\ &\leq \max_{(\upsilon,\omega)\in E(P)} \sum_{\Gamma(\upsilon,\omega)} 2n\frac{P'(\sigma,\beta)}{P(\upsilon,\omega)} \leq \max_{(\upsilon,\omega)\in E(P)} \sum_{\Gamma(\upsilon,\omega)} 2n\frac{1/\binom{n}{2}}{\frac{\lambda}{(1+\lambda)(n-1)}} = O(n^2). \end{aligned}$$

If, on the other hand, $\pi(\upsilon) \geq \pi(\beta)$, then we use detailed balance to obtain:

$$\begin{aligned}
A &= \max_{(\upsilon,\omega)\in E(P)} \left\{ \frac{1}{\pi(\upsilon)P(\upsilon,\omega)} \sum_{\Gamma(\upsilon,\omega)} |\gamma_{\sigma\beta}|\pi(\sigma)P'(\sigma,\beta) \right\} \\
&= \max_{(\upsilon,\omega)\in E(P)} \left\{ \frac{1}{\pi(\upsilon)P(\upsilon,\omega)} \sum_{\Gamma(\upsilon,\omega)} |\gamma_{\sigma\beta}|\pi(\beta)P'(\beta,\sigma) \right\} \\
&\leq \max_{(\upsilon,\omega)\in E(P)} \sum_{\Gamma(\upsilon,\omega)} 2n \frac{P'(\beta,\sigma)}{P(\upsilon,\omega)} \\
&\leq \max_{(\upsilon,\omega)\in E(P)} \sum_{\Gamma(\upsilon,\omega)} 2n \frac{1/\binom{n}{2}}{\frac{\lambda}{(1+\lambda)(n-1)}} \;=\; O(n^2).
\end{aligned}$$

In either case, we have $A = O(n^2)$. Let $\lambda = \min_{i<j} \lambda_{j,i}$. Then $\pi_* = \min_{\rho\in\Omega} \pi(\rho) \geq \lambda^{\binom{n}{2}}/n!$, so $\log(1/(\epsilon\pi_*)) = O(n^2 \log \epsilon^{-1})$, as above. Applying Theorem 2.2.6 proves that $\tau_{nn}(\epsilon) = O(n^9 \log(n/\epsilon))$. □

Remark 4.5.3. *By repeating Stage 1 of the path a constant number of times, it is possible to relax the weakly regular condition slightly if we are satisfied with a polynomial bound on the mixing time.*

4.6 Slow mixing of $\mathcal{M}_{nn}$

We conclude by showing that while $\mathcal{M}_{nn}$ is rapidly mixing for two large, interesting classes of inputs, this is not true in general. In particular, we show that there are positively biased permutations for which the chain $\mathcal{M}_{nn}$ requires exponential time to converge to equilibrium. This disproves the conjecture that the chain will always be fast when $\mathbf{P}$ satisfies $p_{ij} \geq 1/2$ for all $i < j$.

Our example comes from sampling staircase walks with fluctuating bias, which were examined in Chapter 3. Recall that staircase walks are sequences of n ones and n zeros, which correspond to paths from $(0, n)$ to $(n, 0)$, where each 1 represents a step to the right and each zero represents a step down (see Figure 14b). We will show that for our choice of $\mathbf{P}$, permutations are equivalent to staircase walks, and hence

the proof that the Markov chain on staircase walks is slow applies in our setting as well.

Suppose, for ease of notation, that we are sampling permutations with $2n$ entries (having an odd number of elements will not cause qualitatively different behavior). Let $M = 2n^{2/3}$, $0 < \delta < \frac{1}{2}$ be a constant, and $\epsilon = 1/n^2$. For $i < j \leq n$ or $n < i < j$, $p_{i,j} = 1$, ensuring that once the elements $1, 2, \ldots, n$ get in order, they stay in order (and similarly for the elements $n+1, n+2, \ldots, 2n$). The $p_{i,j}$ values for $i \leq n < j$ are defined as follows (see Figure 14a):

$$p_{i,j} = \begin{cases} 1 - \delta & \text{if } i - j + 2n + 1 \geq n + M; \\ \frac{1}{2} + \epsilon & \text{otherwise.} \end{cases}$$

Since the smallest (largest) n elements of the biased permutation never change order once they get put into increasing order, permutations with these elements out of order have zero stationary probability. Hence we can represent the smallest n numbers as ones and the largest n numbers as zeros, assuming that within each class the elements are in increasing order. Given a permutation σ, let $f(\sigma)$ be the sequence of ones and zeros such that $f(\sigma)_i = 1$ if $i \leq n$ and 0 otherwise. Then if σ is such that elements $1, 2, \ldots, n$ and elements $n+1, n+2, \ldots, 2n$ are each in order, $f(\sigma)$ maps σ uniquely to a staircase walk. For example, the permutation $\sigma = (5, 1, 7, 8, 4, 3, 6, 2)$ maps to $f(\sigma) = (0, 1, 0, 0, 1, 1, 0, 1)$. The probability that an adjacent 1 and a 0 swap in $\mathcal{M}_{nn}$ then depends on how many ones and zeros occur before that point in the permutation. Specifically, if element i is a 0 and element $i+1$ is a 1 then we swap them with probability $\frac{1}{2} + \epsilon$ if the number of ones occurring before position x plus the number of zeros occurring after $i+1$ is less than $n + M - 1$. Otherwise, they swap with probability $1 - \delta$. Equivalently, the probability of adding a unit square at position $v = (x, y)$ (i.e. the *bias at* $v = (x, y)$) is $\frac{1}{2} + \epsilon$ if $x + y \leq n + M$, and $1 - \delta$ otherwise; see Figure 14a. This is precisely the case analyzed in Section 3.5.3. Therefore, in this case, the Markov chain $\mathcal{M}_{nn}$ is slow.

Theorem 4.6.1. *There exists a set* $\mathbf{P}$ *for which* $\mathcal{M}_{nn}$ *has mixing time* $\tau(\epsilon) \geq e^{n^{1/3}-1}/4 - 1/2$.

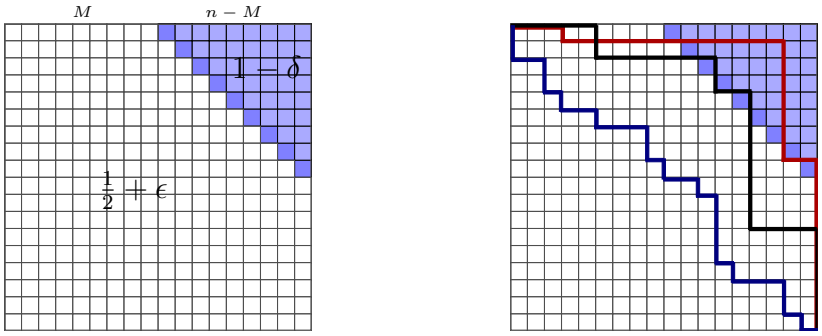


Figure 14: (a) Fluctuating bias with exponential mixing time. (b) Staircase walks in S_1, S_2, and S_3.

In fact, the same proof can be extended to the more general Markov chain where we can swap any 1 with any 0, as long as we maintain the correct stationary distribution. This is easy to see, because any move that swaps a single 1 with a single 0 can only change the maximum height by at most 2 (see Figure 15). If we expand S_2 to include all configurations with maximum height $n+M$ or $n+M+1$, $\pi(S_2)$ is still exponentially smaller than $\pi(S_1) \leq \pi(S_3)$. Hence the Markov chain that swaps an arbitrary pair of elements of the permutations can still take exponential time to converge.

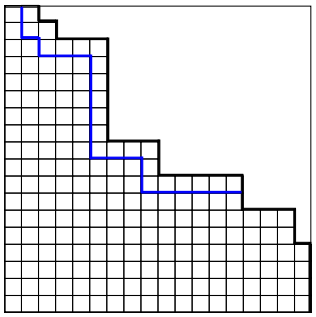


Figure 15: A move that swaps an arbitrary $(1, 0)$ pair.

4.7 Conclusion

The problem of sampling weighted permutations under a positively biased distribution is more challenging than it first appears. The simplicity of our arguments belies

the difficulty of the general problem. We hope that the slow mixing example from Section 4.6 demonstrates the fact that surprising things can happen in the general case. There is much more to be explored in this problem.

One interesting open question is whether the regularity condition proposed by Jim Fill is enough to guarantee that $\mathcal{M}_{nn}$ is rapidly mixing. Our slow-mixing example does not satisfy these conditions. This condition arises from the application to self-organizing lists and is also necessary for our proof of fast mixing for the class of distributions with tree structure. Since the same conditions arose in very different settings, this suggests that they are fundamental is some way. It is likely that none of the techniques developed so far will work in the general setting of regular distributions, so it is an exciting area of future research. A particularly interesting case of the regular distributions is when $p_{i,j} = r_{j-i}$ for $i < j$, so that the probability of swapping i and j depends on the difference between the two.

Another possible direction is to consider in more detail a representation of permutations as linear extensions of a partial order. Bubley and Dyer [12] analyzed $\mathcal{M}_{nn}$ in the context of sampling linear extensions of a partial order, in which case some transpositions are forbidden by setting some of the $p_{i,j}$ values equal to 1. We now understand that the biased surfaces from Section 3 are a special case of the problem of sampling (biased) linear extensions, and in fact as far as we know, it is the only case that is known outside of the uniform distribution. There are many other natural instances of this problem; for example, it is not known whether an analog of the Benjamini et al. [5] result holds in this setting, i.e. $p_{i,j}$ is 1 or p for all $i < j$. Or, for another example, can certain properties of a poset guarantee that $\mathcal{M}_{nn}$ will be efficient for sampling its linear extensions?

CHAPTER V

CLUSTERING IN INTERFERING BINARY MIXTURES

Next, we characterize the high and low density phases for a general family of discrete *interfering binary mixtures* by showing that they exhibit a "clustering property" at high density and not at low density. The clustering property states that there will be a region that has very high area, very small perimeter, and high density of one type of molecule. Special cases of interfering binary mixtures include the Ising model at fixed magnetization and independent sets.

5.1 Clustering of colloids

Colloids are mixtures of two types of molecules in suspension where all non-overlapping arrangements are equally likely. When the density of each type of molecule is low, the mixtures are homogeneous and consequently exhibit properties that make them suitable for many industrial applications, including fogs, gels, foods, paints, and photographic emulsions (see, e.g., [8], [37]). In contrast, when the density is high, the two types of molecules separate whereby one type appears to cluster together. Although this behavior is similar to phase transitions that occur in other discrete models, such as the Ising and Potts models, here the two types of molecules do not possess any *enthalpic* forces causing like particles to attract or disparate particles to repel. In contrast, the behavior of colloids is purely *entropic* — the only restriction is a "hard-core" constraint requiring objects to remain in non-overlapping positions, and clustering occurs at high density because the overwhelming majority of configurations in the stationary distribution are believed to exhibit such a separation. While the experimental study of colloids is pervasive in surface chemistry, material science,

physics, and nanotechnology, there has been little rigorous work explaining their behavior. Even running simulations has been challenging because local algorithms will be slow to converge at high density. Dress and Krauth [23] introduced an algorithm to try to overcome this obstacle, but this too was shown to require time exponential in the number of molecules in some cases [57]. Nonetheless, their algorithm seems to be well-behaved in practice, and Buhot and Krauth [13] provided simulations showing strong heuristic evidence of the presence of two distinct phases in colloid models consisting of different sized squares.

Frenkel and Louis [29] studied an interesting discrete model of colloids whose behavior can be related to the Ising model, a standard model of ferromagnetism. Their model consists of mixtures of unit squares in a region of $\mathbb{Z}^2$ and diamonds of area 1/2 that sit on lattice edges (see Fig. 16). They show that this colloid model, which we call **Model 1**, corresponds to an Ising model, where the density of squares fixes the magnetization and the density of diamonds determines the temperature (see Section 2.1). The Ising model at low temperature is known to exhibit clustering of positive spins. In fact the precise limiting shape of the cluster known as the Wulff shape has been extensively studied using sophisticated techniques (see, e.g. [21], or the references therein). **Model 1** then inherits the phase transition arising in the Ising model which shows there will be clustering (of the squares) at high densities [52]. In this paper we study clustering using elementary methods that apply to a large class of natural colloid models. We characterize clustering directly in terms of the parameters arising from the model to distinguish between the high and low phases and understand the role the density of each type of molecule plays.

We consider a class of *interfering binary mixtures*. Let (Λ_A, Λ_B) be a pair of planar (face-transitive) lattices such that a face of Λ_A and a face of Λ_B are either disjoint, intersect at a single vertex, or intersect at a simply-connected region that is isomorphic to a fixed shape s with nonzero area. We assume for simplicity that

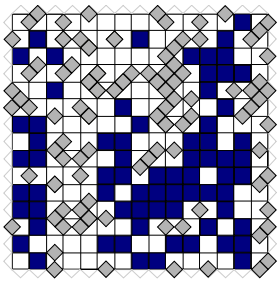


Figure 16: **Model 1**, squares and diamonds on the $n \times n$ grid L_n

Λ_A is one of the following: the grid lattice, the triangular lattice, or the hexagonal lattice; however, the arguments outlined here likely apply in other settings as well. For example, in **Model 1**, Λ_A is the Cartesian lattice $\mathbb{Z}^2$ and Λ_B is the set of diamonds bisected by edges in $\mathbb{Z}^2$; then s is an isosceles triangle with unit base and height $1/2$ (Fig. 16). We consider the intersection of these lattices with some finite region L, where $L_A = \Lambda_A \cap L$ and $L_B = \Lambda_B \cap L$. We are given a set of tiles; A-tiles lie on the faces of L_A and B-tiles lie on the faces of L_B with the additional requirement that tiles must not overlap. In Section 5.5, we will give examples of other interfering binary mixtures, including independent sets, that arise naturally in combinatorics and statistical physics and contrast these with a non-interfering binary mixture that provably does not exhibit clustering.

It is often useful to switch from a model where the number of tiles of each type are fixed to a so-called *grand-canonical ensemble* where these are allowed to vary. Here, however, typical configurations would have a preponderance of only one type of tile at most high densities and the balanced configurations we are interested in would be exponentially unlikely. Instead, we fix the number of A-tiles and allow the B-tiles to vary stochastically. Each configuration σ has weight proportional to $\lambda^{d(\sigma)}$, where $d(\sigma)$ is the number of B-tiles in σ. The choice of λ controls the expected density of B-tiles.

Our goal now is to understand when the A-tiles will exhibit clustering in terms of the (expected) density of A-tiles and B-tiles. First we define a clustering property for

configurations of tiles. Informally we have clustering if there exists a dense region R in Λ_A with $\Omega(n^2)$ area and $O(n)$ perimeter. Our main theorems demonstrate that at high density interfering binary mixtures exhibit the clustering property while at low densities they do not. We give precise definitions of the clustering property and state the main theorems in Section 5.2. In Sections 5.3 and 5.4 we prove the two main theorems in the context of **Model 1** and in Section 5.5 we explain the generalization to other interfering binary mixtures.

The key tools in our proofs are careful *Peierls arguments*, used in statistical physics to study uniqueness of the Gibbs state and phase transitions (see, e.g., [20], [22]), and in computer science to study slow mixing of Markov chains (see, e.g., [10], [32], [61]). Peierls arguments allow you to add and remove contours by complementing the interiors of those contours. The main challenge here is maintaining the number of A-tiles, making the arguments considerably more difficult. We introduce the concept of bridge systems, to handle multiple contours by connecting components and to make it possible to efficiently encode the boundaries of all contours removed. The encoding is necessary to account for the entropy/energy tradeoffs in these maps.

5.2 Binary mixtures and the clustering property

We begin by formalizing the model, defining clustering and stating our main theorems.

5.2.1 Interfering binary mixtures

Recall A-tiles lie on faces of $L_A = \Lambda_A \cap L$ and let $|L_A|$ be the total number of faces of L_A. Given constants $\lambda > 1$, and $0 < b < 1/2$, where $b|L_A| \in \mathbb{Z}$, define $\Omega = \Omega(b, \lambda)$ as the set of non-overlapping packings of L with $b|L_A|$ A-tiles and any number of B-tiles (where a tile can only be placed on a face of its type). We wish to study the distribution $\pi(\rho) = \lambda^{d(\rho)}/Z$, where $d(\rho)$ is the number of B-tiles in ρ and $Z = \sum_{\rho \in \Omega} \lambda^{d(\rho)}$ is a normalizing constant. Our goal is to determine whether a

configuration chosen according to π is likely to have clusters of A-tiles.

In Sections 5.2 - 5.4, we study **Model 1**; these techniques can be extended to other models of interfering binary mixtures, which we will demonstrate in Section 5.5. As mentioned earlier, **Model 1** is equivalent to the Ising model of ferromagnetism with a fixed magnetization, which we will see presently. First, we will define the Ising model on the $n \times n$ grid L_n. Let $\bar{G} = (\bar{V}, \bar{E})$ be the dual lattice region to L_n and let $\rho \in \{+, -\}^{\bar{V}}$ be an assignment of spins to each of the vertices in $\bar{V}$ (i.e., the faces in V). The weight of a configuration is $\bar{\pi}(\rho) = e^{\beta|\bar{E}_d(\rho)|}/\bar{Z}$, where $\bar{E}_d(\rho) \subseteq \bar{E}$ is the set of edges in $\bar{G}$ whose endpoints have different spins in ρ, β is the inverse temperature and $\bar{Z}$ is a normalizing constant.

Returning to **Model 1**, let the *square structure* $\Gamma(\rho)$ of a configuration ρ in Ω be the configuration σ obtained from ρ by removing all of its B-tiles (diamonds). The set $\widehat{\Omega}$ of all such square structures with bn^2 A-tiles (squares) is called the *projection of* **Model 1**. Let $\widehat{\pi}$ be the induced distribution on $\widehat{\Omega}$; that is, for $\sigma \in \widehat{\Omega}$, let $\widehat{\pi}(\sigma) = \sum_{\rho \in \Gamma^{-1}(\sigma)} \pi(\rho)$. Then the function $f : \widehat{\Omega} \to \{+, -\}^{\bar{V}}$, which replaces each square by a positive spin and each empty face by a negative spin, is a bijection which maps the projection of **Model 1** onto the Ising model. To see this, define the *perimeter* of σ (in Ω or $\hat{\Omega}$) to be the edges that belong to exactly one A-tile in σ, and define $\kappa(\sigma)$ as the length of the perimeter of σ. Let $e(\sigma)$ be the number of edges that are not incident to any A-tile in σ. We find that

$$\widehat{\pi}(\sigma) = \sum_{k=0}^{e(\sigma)} \frac{\lambda^k}{Z} \binom{e(\sigma)}{k} = \frac{1}{Z}(1+\lambda)^{e(\sigma)} = (1+\lambda)^{2n^2-2bn^2} \frac{\mu^{\kappa(\sigma)}}{Z}, \tag{9}$$

where $\mu = (1+\lambda)^{-\frac{1}{2}}$. Thus, the total perimeter of the square structure completely determines the probability that it will show up in Ω. Since the weight of a configuration is determined exactly by the number of edges with opposite spins in L_n, this is the Ising model with a fixed number of positive spins for some λ that is a function of β, known as *fixed magnetization*.

While the perimeter of the square structure does not exactly determine its probability for the other models of interfering binary mixtures, we will see that they are closely related, and we can still use arguments about the perimeter to bound the weight of configurations. Thus it makes sense to define the clustering property in terms of the perimeter to area ratio, which we do next.

5.2.2 The clustering property

The goal of this paper is to show that when the density of B-tiles is high, interfering binary mixtures cluster, while at low density they do not. First, we characterize clustering in this context. Intuitively, a configuration has the clustering property if there is a large region densely filled with A-tiles. More precisely, let a *region* R be any set of faces in L_n. The perimeter $\kappa(R)$ of a region R is the number of edges adjacent to a face in R and a face in $\bar{R} = L_n \setminus R$. Let $c = \min\left\{\frac{b}{2}, \frac{1}{100}\right\}$.

Definition 1. *We say that a configuration $\sigma \in \Omega$ (or $\Gamma(\sigma) \in \widehat{\Omega}$) has the* clustering property *if it contains a region R which satisfies the following properties:*

1. *R contains at least $(b-c)n^2$ A-tiles,*
2. *the perimeter of R is at most $8\sqrt{b}\ n$, and*
3. *the density of A-tiles in R is at least $1-c$ and in $\bar{R}$ is at most c.*

If a configuration has the clustering property, we show that it contains an $n^{1/3} \times n^{1/3}$ window with high density and one with low density, demonstrating the heterogeneity of the configuration. In Section 5.5.1 we contrast this with **Model 2**, related to bond percolation, which remains homogeneous at all densities.

5.2.3 Main results

We show that at high density interfering binary mixtures have the clustering property while at low densities they do not. Specifically, we prove the following theorems in

the context of **Model 1** on the $n \times n$ region L_n with bn^2 A-tiles and the density of B-tiles determined by λ. In Section 5.5, we show they also hold for other interfering binary mixtures.

Theorem 5.2.1. *For $0 < b \leq 1/2$, there exist constants $\lambda^* = \lambda^*(b) > 1, \gamma_1 < 1$ and $n_1 = n_1(b)$ such that for all $n > n_1$, $\lambda \geq \lambda^*$ a random sample from Ω will have the clustering property with probability at least $(1 - {\gamma_1}^n)$.*

Theorem 5.2.2. *For $0 < b < 1/2$, there exist constants $\lambda_* = \lambda_*(b) > 0, \gamma_2 < 1$ and $n_2 = n_2(b)$ such that for all $n > n_2$, $\lambda \leq \lambda_*$ a random sample from Ω will not have the clustering property with probability at least $(1 - {\gamma_2}^n)$.*

Furthermore, it follows from the proofs that at low density if a dense region R' has area $\Omega(n^2)$ then it must have perimeter $\Omega(n^2)$. Notice that in the case $b > 1/2$ we can obtain comparable results by the symmetry of the A-tiles to the empty space. Indeed, in this case if λ is sufficiently high we will see empty cells clustering within a sea of A-tiles and for low density the empty cells will be well-distributed.

Note that since clustering is just a property of the A-tiles, it suffices to prove Theorems 5.2.1 and 5.2.2 for weighted square structures $\widehat{\Omega}$, involving just the A-tiles. From this point we focus on $\widehat{\Omega}$, and we refer to A-tiles just as *tiles*.

5.3 High density of B-tiles

We concentrate first on interfering binary mixtures at high density to prove Theorem 5.2.1. Define $\Psi \subset \hat{\Omega}$ to be the set of configurations that have the clustering property; then we show that $\widehat{\pi}(\hat{\Omega} \setminus \Psi) \leq \gamma_1^n \widehat{\pi}(\Psi)$ for some constant $\gamma_1 < 1$. To achieve this, we apply a Peierls argument, in which we define a map $f : \hat{\Omega} \setminus \Psi \to \Psi$ and show that for all $\tau \in \Psi$,

$$\sum_{\sigma \in f^{-1}(\tau)} \widehat{\pi}(\sigma) \leq \gamma_1^n \widehat{\pi}(\tau). \tag{10}$$

Given a configuration $\sigma \in \hat{\Omega} \setminus \Psi$, the map f removes a large set T of tiles in σ and reassembles them in a single large component in $f(\sigma)$. This decreases the total perimeter of the configuration significantly, and therefore $\widehat{\pi}(f(\sigma))$ is exponentially larger than $\widehat{\pi}(\sigma)$. The challenge is to bound the number of configurations that map to a given $\tau \in \Psi$ by carefully encoding the preimages of τ.

Some definitions will be helpful. We say two tiles are *adjacent* if their borders share an edge. A *component* is a maximal connected set of tiles, and maximal connected segments of the perimeter of σ are *contours*. The set T of tiles we remove will be a union of components, which we identify using a system of "bridges" connecting these components (Fig. 17). The key is that the number of edges in the bridges is at most a constant times the total perimeter of the components bridged. Then if E is the set of all edges in bridges or along contours bridged, we can bound $|f^{-1}(\tau)|$ by the number of ways that those E edges could be distributed in σ. Finally, we show that there is a sparse, roughly square region in the resulting configuration where we can add the T tiles. We complement that region to obtain $f(\sigma)$, which allows us to remember the locations of any components that were not bridged (see Figure 17). Notice that the resulting configuration has much higher weight (much smaller perimeter), as desired.

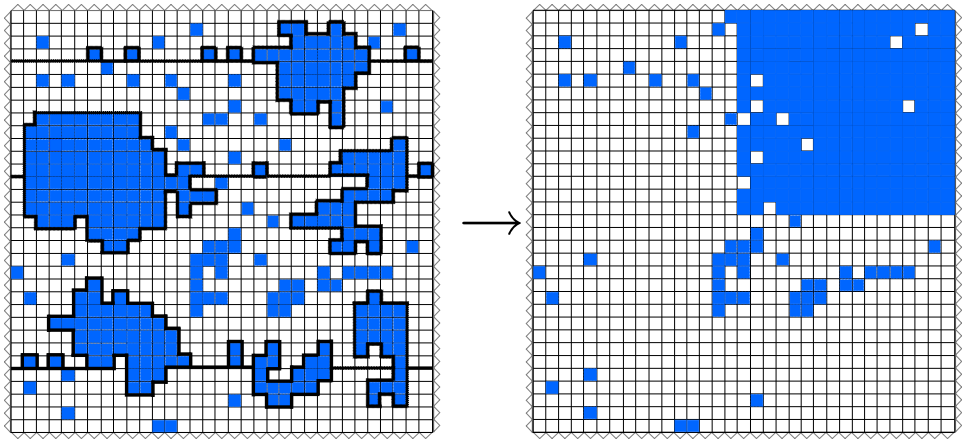


Figure 17: A configuration $\sigma \in \widehat{\Omega} \setminus \Psi$ and the image $f(\sigma)$ of σ in Ψ

Building Bridges. Given a region R, let $C(R)$ be the set of contours fully contained

within the interior of R and define the *outer contours* to be those in $C(R)$ that are not contained in the interior of other contours in $C(R)$. The interior of the outer contours of components are called *holes* and the interior of the outer contours of holes are called *islands*.

Consider first the case in which there are no components with holes. Suppose B is a set of edges of L_n connecting some subset S of the contours to the boundary of L_n. We call B a set of *bridges* and S a set of *bridged contours*. A cell in L_n or a tile is called *unbridged* if it is not bounded by a bridged contour. Then (B, S) is a **c-bridge system** for $\sigma \in \widehat{\Omega}$ if the number of unbridged tiles is at most c times the number of unbridged cells, and $|B| \leq \kappa(S)(1-c)/(2c)$. If σ has components with holes, then first construct a c-bridge system (B, S) for σ', obtained from σ by filling all the holes. Next for each bridged contour X in σ, construct a c-bridge system for the region in σ bounded by X (treating tiles as empty cells and empty cells as tiles). Recurse until you obtain c-bridge systems for each bridged contour at every level of the recursion. We call this a c-bridge system of σ.

Lemma 5.3.1. *There exists a c-bridge system for any configuration $\sigma \in \hat{\Omega}$.*

Proof. If any components of σ have holes, we may need to recurse as described above. We may assume that we are given a region R in σ with no holes, since otherwise we recurse as described above. Now we use induction on the number of contours in R. If there are no contours, then clearly $(\emptyset, \emptyset)$ is a c-bridge system for R. Otherwise, define $t(R)$ to be the set of tiles in R and $x(R)$ to be the number of empty cells in R. Let $\mathbb{H}$ be the set of horizontal lines through R. If, for every $H \in \mathbb{H}$, $|t(R) \cap H| < c|R \cap H|$ then we are done, since then $(\emptyset, \emptyset)$ is a $c-$bridge system for R. Otherwise there exists a horizontal line H such that $|t(R) \cap H| \geq c|R \cap H|$. Then let B be the set of bottom edges of every outer cell in $H \cap R$. See Fig. 18, where the dark black edges along the line H are the new bridges. Let S be the set of contours connected in this step. We know that $\kappa(S) \geq 2|t(R) \cap H| \geq 2c\ |R \cap H| \geq 2c/(1-c)|x(R) \cap H|$, so

$|B| \leq (1-c)/(2c)\kappa(S)$. We obtain R' from R by removing the cells bounded by a contour in S, as in Fig. 18. Then by induction, there exists a $c-$bridge system (B', S') of R'. Then $\widehat{B} := B \cup B'$ is a set of bridges connecting the contours in $\widehat{S} = S \cup S'$ to each other and to the boundary of R. Moreover, $|\widehat{B}| \leq \frac{1-c}{2c}\ \kappa(\widehat{S})$ and the number of unbridged tiles is at most c times the number of unbridged cells. Hence $(\widehat{B}, \widehat{S})$ is a c-bridge system for R. □

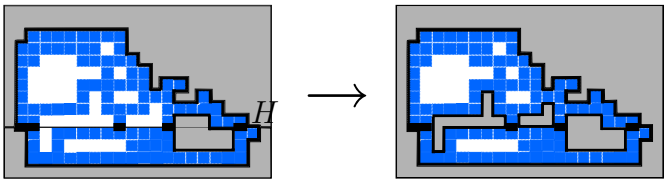


Figure 18: Before and after one step of the construction of a c-bridge system for a region R; the solid lighter grey area is exterior to R

Once we have a c-bridge system, we can apply a map in which we *complement* an entire region of cells, making tiled cells empty and vice versa. This map significantly reduces the perimeter, but can dramatically change the total number of tiles. Recall we must maintain the total number of tiles, so we may need to supplement by adding extra tiles from another region or we may have extra tiles, which we will put in our "bank" for later. At the end of the process we will find a roughly square region that we can again complement using the bank of extra tiles so that the total number of tiles is restored to bn^2 at minimal cost.

Finding a Sparse Box. We now show that after removing all but cn^2 tiles, there exists a roughly square region of low density where we can place the tiles in our bank.

Lemma 5.3.2. *For $(b-c)n^2 \leq a < bn^2$, there exists a constant $n_3 = n_3(b)$ such that for all $n \geq n_3$, if ρ is a configuration with at most cn^2 tiles then ρ contains a roughly square region R' such that complementing R' requires a additional tiles and the change in total perimeter is at most $5\sqrt{a}$.*

Proof. Given a region R, let $d(R)$ denote the number of tiles needed to complement R; this is exactly the area of R minus twice the number of tiles in R. Let $l = \lceil \sqrt{8a/7} \rceil$. First we show that there exists a square $l \times l$ region R such that $d(R) \geq a$. Assume that such a square does not exist. Divide the grid into $\left\lfloor \frac{n}{l} \right\rfloor^2$ disjoint squares with side length l and consider any square. Let t be the number of tiles in the square. The empty volume is at least $l^2 - t$. By assumption each square satisfies $l^2 - t < t + a$, and so $t > \frac{l^2 - a}{2}$. In particular, $8a/7 \leq l^2 < a + 2t \leq a + 2cn^2$, so we know $a < 14cn^2$. This implies that $l \leq \sqrt{8a/7} + 1 \leq 1 + 4\sqrt{c}n$. However, if T is the total number of tiles,

$$cn^2 \geq T > \left\lfloor \frac{n}{l} \right\rfloor^2 \frac{l^2 - a}{2} \geq \frac{n^2}{2} \left(1 - \frac{l}{n}\right)^2 \left(1 - \frac{a}{l^2}\right) > \frac{n^2 \left(1 - \frac{1}{n} - 4\sqrt{c}\right)^2}{16} \geq cn^2,$$

since $c \leq \frac{1}{65}$ and $n \geq n_3$, a contradiction. Therefore there exists an $l \times l$ square R such that $d(R) \geq a$. Remove cells from R one at a time, starting with the bottom row of R and moving across, until we obtain a region $R' \subseteq R$ with $d(R') = a$. This can be done because removing one cell at a time changes d by at most 1. This region R' is roughly square and has perimeter at most $4\sqrt{8a/7} < 5\sqrt{a}$.

□

The Proof of Theorem 5.2.1. Finally we can prove Theorem 5.2.1, showing that for large λ a typical configuration will have the clustering property.

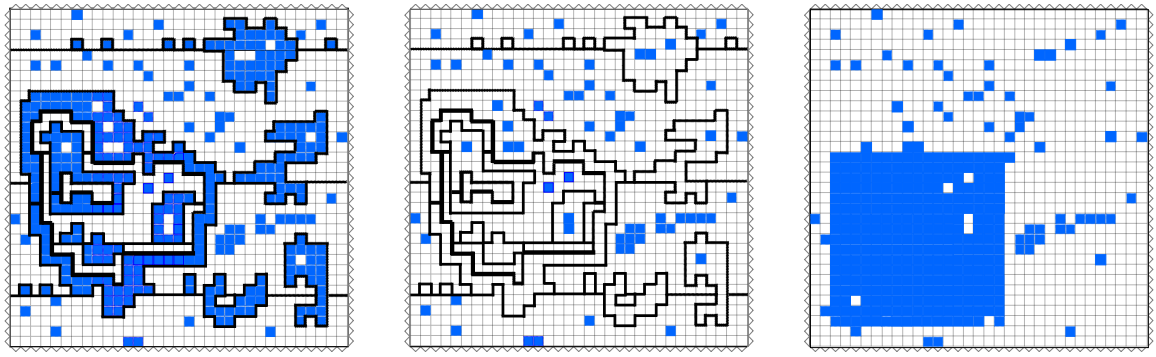


Figure 19: A c-bridge system for $\sigma \in \hat{\Omega} \setminus \Psi$; the image $f_1(\sigma)$; and $f(\sigma) = f_2 \circ f_1(\sigma)$

Proof of Theorem 5.2.1. Let $\sigma \in \hat{\Omega} \setminus \Psi$. Construct a c-bridge system (B, S) for L_n as described in Lemma 5.3.1. That is, (B, S) is a set of bridges in L_n connecting some of the components, some of the holes within those components, some of the islands within those holes, etc. For any bridged contour X, let $r(X)$ be the region bounded by X. If $r(X)$ is a component with holes, then we remove all outer tiles of $r(X)$ and complement all unbridged holes in X, using a subset of the tiles removed to fill in the holes. If $r(X)$ is a hole with islands, then we leave all of the unbridged islands alone. At this point, after complementing some number of regions, we have a bank of extra tiles; let a be the number of tiles in the bank. Notice that by the definition of a c-bridge system, the density of tiles remaining is at most c, so $a \geq (b-c)n^2$.

Let $f_1(\sigma)$ be obtained from σ by removing the bridged components and complementing as described above. Let $\mathcal{F}_1$ be the image of f_1 on $\hat{\Omega}\setminus\Psi$; note that $\mathcal{F}_1 \not\subset \hat{\Omega}$ since the configurations in $\mathcal{F}_1$ have too few tiles. Let κ be the total perimeter of all contours bridged. Then for any $\rho \in \mathcal{F}_1$, we claim that the number of preimages of ρ whose bridged contours have total perimeter κ is at most 5^{c_3} for $c_3 = (1 + \frac{1-c}{2c} + \frac{1}{8\sqrt{a}})\kappa$. Consider the c-bridge system obtained above for L_n. Let V denote the leftmost vertical edges of the region. Let $S' = S \cup V$. We perform what is essentially a depth-first-search traversal of the bridge system on S', starting at the top left corner of L_n. As we traverse an edge we record what type of edge it was using. Then we 'encode' the location of the bridges and contours using five bits that represent forward, left, right, bridge east, or bridge west; note that all bridges are horizontal edges, so all edges in B fall into one of these 5 categories. Whenever we encounter a new bridge B_i, we "process" that bridge by traversing it from the previous contour C_i to the next contour C_{i+1}, then traversing the edges of C_{i+1}. If we encounter another bridge, we process it before continuing. We finish processing B_i when we return to the intersection of C_{i+1} with B_i. Finally, we jump back to the intersection of B_i with C_i and

continue traversing C_i. Given the encoded information, there is a unique way to distribute the contours. Hence for all perimeters $\kappa \geq 8\sqrt{a}n$ the number of preimages of ρ whose bridged contours have total perimeter κ is at most $5^{|B|+\kappa+n} \leq 5^{c_3}$. Therefore $|f_1^{-1}(\rho)| \leq \sum_{\kappa \geq 8\sqrt{a}n} 5^{c_3}$.

Let $\rho \in \mathcal{F}_1$ with $bn^2 - a$ tiles. Lemma 5.3.2 shows how to find a region S' in ρ to complement using the a tiles from the bank to obtain τ in such a way that $\kappa(\tau) - \kappa(\rho) \leq 5\sqrt{a}$. Let $f_2(\rho) = \tau$ and $f = f_2 \circ f_1$. We can encode the boundary of S' with $n^2 3^{\kappa(S')} \leq n^2 3^{5\sqrt{a}}$ information. Hence for any $\tau \in \Psi$,

$$|f^{-1}(\tau)| \leq n^2 3^{5\sqrt{a}} \max_{\rho \in f_2^{-1}(\tau)} |f_1^{-1}(\rho)|.$$

Let $\sigma \in \hat{\Omega} \setminus \Psi$, and as above let κ be the total perimeter of components bridged in σ (recall $\kappa(\sigma)$ is the total perimeter of all contours in σ). If $\kappa \leq 8\sqrt{a}$, then $\sigma \in \Psi$, a contradiction. To see this, define the parity of a cell to be 1 if it is contained within an odd number of bridged contours and 0 otherwise, and let R be the set of cells with parity 1. Then R has density at least $1 - c$, perimeter at most $8\sqrt{a}$ and $a \geq (b-c)n^2$ tiles. Moreover, $\bar{R}$ has density at most c. Thus R is the region we require, and so $\sigma \in \Psi$. This implies $\kappa > 8\sqrt{a}$. We have shown that $\kappa(\sigma) - \kappa(f(\sigma)) > \kappa - 5\sqrt{a} > \kappa/4$. Let $\tau \in \Psi$ and define $f_\kappa^{-1}(\tau)$ to be the set of configurations with perimeter κ that map to τ. Then $|f_\kappa^{-1}(\tau)| \leq n^2 \left(3^{\frac{1}{2\sqrt{7}}} 5^{1+\frac{1-c}{2c}+\frac{1}{16\sqrt{b}}}\right)^\kappa$ and so

$$\pi(\tau)^{-1} \sum_{\sigma \in f^{-1}(\tau)} \pi(\sigma) \leq \sum_{\sigma \in f^{-1}(\tau)} \mu^{\kappa(\sigma)-\kappa(f(\sigma))} \leq \sum_{\kappa=8\sqrt{a}}^{2n^2} \mu^{\kappa/4} |f_\kappa^{-1}(\tau)| \leq \gamma_1^n,$$

for some $\gamma_1 < 1$, if $\mu \leq \mu^* < \left(3^{\frac{1}{2\sqrt{7}}} 5^{1+\frac{1-c}{2c}+\frac{1}{16\sqrt{b}}}\right)^{-4}$. Thus the theorem holds if $\lambda \geq \lambda^* = \mu^{*-2} - 1$. □

As a corollary, we find that if a configuration has the clustering property then there exists an $n^{1/3} \times n^{1/3}$ window with high density and one with low density.

Corollary 1. *For $0 < b \leq 1/2$ there exists a constant $n_4 = n_4(b)$ such that for all $n > n_4$, if σ satisfies the clustering property then σ contains square $n^{1/3} \times n^{1/3}$*

windows W_1 and W_2 such that the density of tiles in W_1 is at least $.99(1-c)$ and the density of tiles in W_2 is at most $2.1c$.

Proof. Let $\sigma \in \Psi$ and let R be the active region given by the clustering property. Consider the set of $n^{4/3}$ windows of side length $n^{1/3}$ that tile L_n. Since R has $a \geq (b-c)n^2$ tiles and density at least $1-c$, we know that at most $an^{-2/3}/(1-c)$ windows are contained completely within R. Similarly, since R has perimeter at most $8\sqrt{b}\ n$ we know that at most $8\sqrt{b}\ n$ windows intersect the boundary of R. This means that there exists a constant $n_5 = n_5(b)$ such that for $n > n_5$, there exists a window with density d satisfying,

$$d \geq \frac{a}{n^{2/3}\left(\left(\frac{a}{1-c}\right)n^{-2/3} + 8\sqrt{b}\ n\right)} \geq \frac{99}{100}(1-c).$$

Next, consider the region $\bar{R} = L_n - R$. From the clustering property we know that $\bar{R}$ has area at least $n^2 - a/(1-c)$ and contains at most cn^2 tiles. This implies that there are at least $n^{-2/3}(n^2 - a/(1-c))$ windows intersecting $\bar{R}$. At most $8\sqrt{b}\ n$ of these windows can intersect R and these contain at most $8\sqrt{b}\ n^{5/3}$ tiles from R. Combining these observations implies that there exits a constant $n_6 = n_6(b)$ such that for $n > n_5$, there exists a window with density d satisfying,

$$d \leq \frac{cn^2 + 8\sqrt{b}n^{5/3}}{n^{2/3}n^{-2/3}\left(n^2 - \frac{a}{1-c}\right)} \leq \frac{99}{49}c.$$

□

5.4 Low density of B-tiles

We now examine the low density case and prove Theorem 5.2.2, stating that typical configurations will not have the clustering property. For small enough λ, the A-tiles will be well-distributed throughout L_n, in the following sense. Any large dense region must have perimeter on the order of n^2.

Proof of Theorem 5.2.2. Define $\delta = ((1-c)/(b-c))^{b-c}$. Let $\Psi' \subset \widehat{\Omega}$ be the set of configurations with a region R that have density at least $1-c$, at least $(b-c)n^2$ tiles,

and perimeter less than αn^2, where α satisfies $0 < \alpha < (\ln(\delta) - b\ln 2)/((1+1/c)\ln 5)$. We will show $\widehat{\pi}(\Psi')$ is exponentially small. Clearly $\Psi \subset \Psi'$, so this implies that the clustering property is exponentially unlikely to occur.

For each $\sigma \in \Psi'$, construct a c-bridge system for σ. As in the proof of Theorem 5.2.1, we complement all bridged components and all non-bridged holes within those components. We obtain $f_1(\sigma)$, which has $t_\sigma \leq cn^2$ tiles, and a bank of $a_\sigma \geq (b-c)n^2$ tiles. Next we define $N(\sigma)$ to be the set of all configurations obtained from $f_1(\sigma)$ by adding a_σ tiles back at any empty location; then $|N(\sigma)| = \binom{n^2 - t_\sigma}{a_\sigma}$. For each $\tau \in \widehat{\Omega}$, we need to bound the number of configurations σ such that $\tau \in N(\sigma)$. As before, we can reconstruct the bridge system for σ with $5^{(1+\frac{1-c}{2c})\kappa+n}$ information and we can recover the original with 2^{bn^2} information by recording whether each tile moved. Hence the number of σ that map to τ is at most $5^{(1+\frac{1-c}{2c})\alpha n^2+n} 2^{bn^2} \leq (2^b\delta)^{n^2/2}$.

Finally, we define a weighted bipartite graph $G(\Psi', \widehat{\Omega}, E)$ with an edge of weight $\pi(\sigma)$ between $\sigma \in \Psi'$ and $\tau \in \widehat{\Omega}$ if $\tau \in N(\sigma)$. The total weight of edges is

$$\sum_{\sigma\in\Psi'} \pi(\sigma)|N(\sigma)| \geq \sum_{\sigma\in\Psi'} \pi(\sigma)\binom{n^2 - (bn^2 - a_\sigma)}{a_\sigma} \geq \pi(\Psi')\delta^{-n^2}.$$

However, the weight of the edges is at most $\sum_{\tau\in\widehat{\Omega}} \pi(\tau)\mu^{-4(b-c)n^2}(2^b\delta)^{n^2/2}$. Let $\mu^* = (2^b/\delta)^{1/(8(b-c))}$ and $\lambda^* = (\mu^*)^{-2} - 1$. Thus for all $\mu < \mu^*$,

$$\pi(\Psi') < \mu^{-4(b-c)n^2}(2^b\delta)^{n^2/2}\delta^{-n^2} < \gamma_2^n,$$

for some $\gamma_2 < 1$, completing the proof. □

5.5 Other models

We conclude by considering other natural models of binary mixtures. First we show that noninterfering models may not have clustering by considering a model related to bond percolation in Section 5.5.1. Next, in Section 5.5.2, we give other examples of interfering binary mixtures. In Section 5.5.3, we show how to extend the analysis of Section 5.3 to prove that all interfering models exhibit clustering.

5.5.1 Noninterfering binary mixtures

Model 2: A-tiles are unit squares on L_n and B-tiles are squares of side length $1/2$ on the half-integer lattice, (see Fig. 21(c)). This model is qualitatively different from the previous models since *the placement* of the A-tiles does not influence *the number of places* in which we can put the B-tiles. In fact, this model is just bond percolation on a rotated grid with a fixed number of edges, where we do not expect clustering at any density. To see the bijection, label a unit square with a Northwest-Southeast diagonal if it lies on an even face and label it with a Northeast-Southwest diagonal otherwise, as in Fig. 21(d). Notice that these lines form a subset of the edges of a rotated grid. If we have bn^2 A-tiles then each edge in the rotated grid is present with probability b.

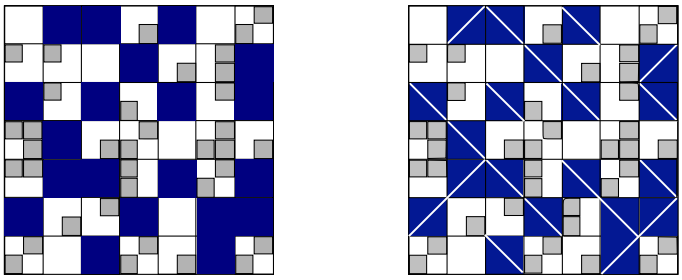


Figure 20: **Model 2** and the connection with bond percolation

To illustrate the difference between the behavior of **Model 2** and the interfering binary mixtures, consider an $n^{1/3} \times n^{1/3}$ window in each. In **Model 2**, the probability that any $n^{1/3} \times n^{1/3}$ box has density d such that $d > 1.5b$ or $d < 0.5b$ is less than γ_3^n for some constant $\gamma_3 < 1$. This is straightforward to show since each configuration of bn^2 tiles in n^2 locations has equal likelihood. Thus, the probablity that a fixed window has density d is exactly $\binom{n^{2/3}}{dn^{2/3}}\binom{n^2-n^{2/3}}{bn^2-dn^{2/3}}/\binom{n^2}{bn^2}$. Using standard approximations and union bounds we can obtain the desired result. In contrast, by Corollary 1, a configuration with the clustering property has a window with density $d \geq .99(1-c)$ and a window with density $d \leq 2.1c$. Hence we see markedly different behavior between interfering

and non-interfering binary mixtures.

5.5.2 Interfering binary mixtures

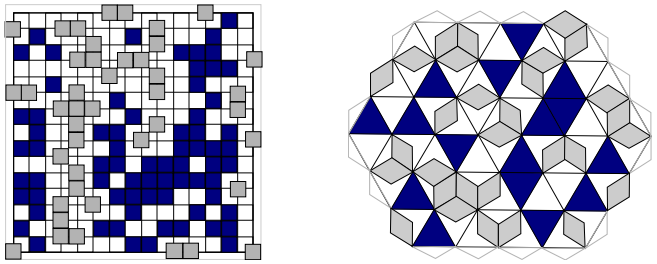
First, we will see a few other examples of interfering binary mixtures, which help to illuminate the slight changes necessary to generalize the proof of Theorem 5.2.1. Then we will describe the whole class.

5.5.2.1 Examples

Model 3: A-tiles are squares on L_n and B-tiles are unit squares centered on vertices of L_n, (see Fig. 21(a)). It is not hard to see that this model corresponds exactly to an independent set model on the rotated grid where vertices correspond to the centers of A-tiles and B-tiles, and the number of even vertices is fixed. The number of odd vertices varies according to λ. Again the A-tiles will cluster together at high enough λ, leaving large regions to fill with B-tiles. The weight of a configuration σ is proportional to λ^v, where v is the number of vertices in σ not intersecting any A-tiles (we call these *open vertices*). Hence we must argue that by removing several components and putting them together into a single large component, the number of open vertices increases. Indeed, the change in the number of open vertices is proportional to the change in the perimeter after this operation. One must be careful, however, to define a component so that two tiles are adjacent if they share a vertex (not an edge). Otherwise, if a region looks like a checkerboard of tiles to empty space, and we remove every other row to create a new component, we decrease the perimeter but increase the number of occupied vertices. This cannot happen as long as we choose maximal connected subsets of tiles according to this definition of adjacency. We describe more details in Section 5.5.3.2.

Model 4: A-tiles are triangles on the triangular lattice Λ_A and B-tiles are lozenges bisected by edges of Λ_A, (see Fig. 21(b)). **Model 4** maps bijectively onto an Ising Model with fixed magnetization on Λ_A. In models like this, where the A-tiles are not

square, the large component we create for Theorem 1 might not be square, but some other roughly rectangular shape such as the one in Figure 25a. The remaining details are similar.



(a) (b)

Figure 21: (a) **Model 3** (b) **Model 4**

5.5.2.2 The class of interfering binary mixtures

Next, we describe the class of interfering binary mixtures. Let the A-lattice be any of the following: the grid, the triangular lattice, or the hexagonal lattice, and let the B-lattice Λ_B be any planar lattice that is face-transitive. Recall that an A-face and a B-face can intersect trivially, or at a vertex, or at a fixed shape s. Since the A-faces (and B-faces) are all isomorphic, this leaves just a few possibilities. Let t_A and t_B be a pair of intersecting tiles on the A and B lattices, respectively. Then the boundary of t_B either intersects the boundary of an A-face at a vertex of t_A or at the midpoint of an edge of t_A. Moreover, the boundary of t_B must intersect the center of t_A. Thus it is easy to check that if e is the number of edges of an A-tile, then the s-shapes can divide the A-tiles into $e/2, e$, or $2e$ parts. All of the possibilities on the square grid are given in Figure 22, where the A-tiles are in blue and the B-tiles are in gray. This includes **Model 1** and **Model 3**, as well as Inverse-Ising, where we swap the roles of the A- and B- lattices, and Half-Ising, where each edge of the grid is in bijection with a *pair* of B-tiles. The possibilities on the other lattices are given in Figures 23 and 24.

These models can all be seen as independent sets on some graph (including the

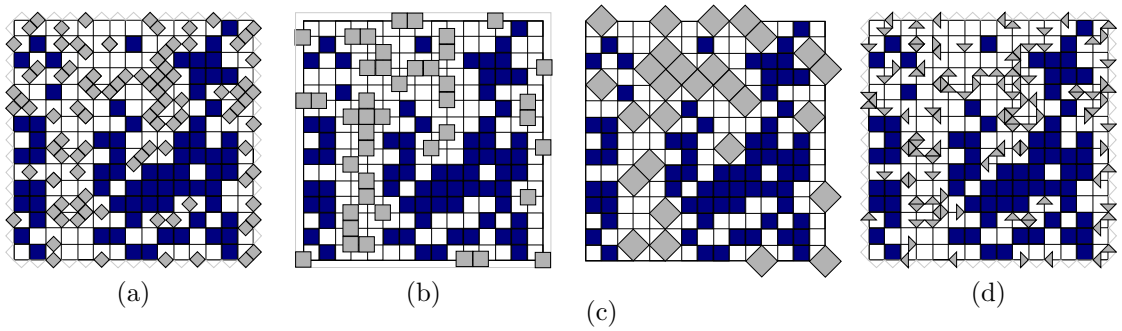


Figure 22: Models on the grid: (a) **Model 1**: Ising, (b) **Model 3**: Independent Sets, (c) Inverse Ising, (d) Half-Ising

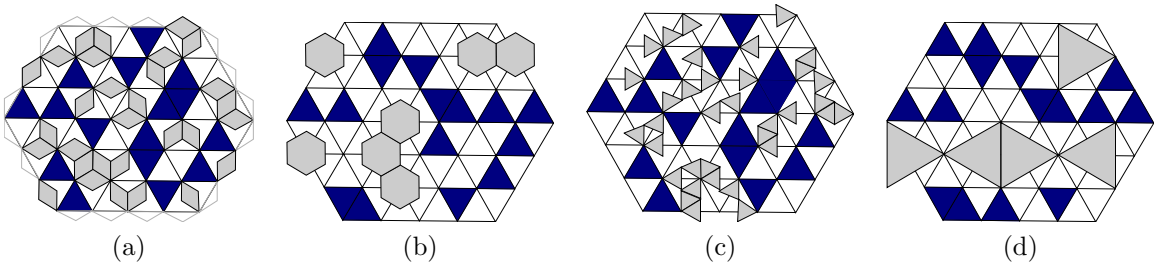


Figure 23: Models on the triangular lattice: (a) **Model 4**: Ising, (b) Independent Sets, (c) Half-Ising, (d) Inverse Half-Ising

Ising model), namely, the graph whose vertex set is the set of centers of all A-faces and B-faces and where two vertices are adjacent if their corresponding faces intersect nontrivially. The weight of a configuration σ of A-tiles is proportional to $\lambda^{\mathcal{O}(v)}$ where $\mathcal{O}(v)$ is the number of *open B-vertices*; that is, B-vertices which do not intersect any A-tiles in σ.

Our goal is to show that our process outlined in Section 5.3 for changing the perimeter of a configuration does in fact change the number of open B-vertices as

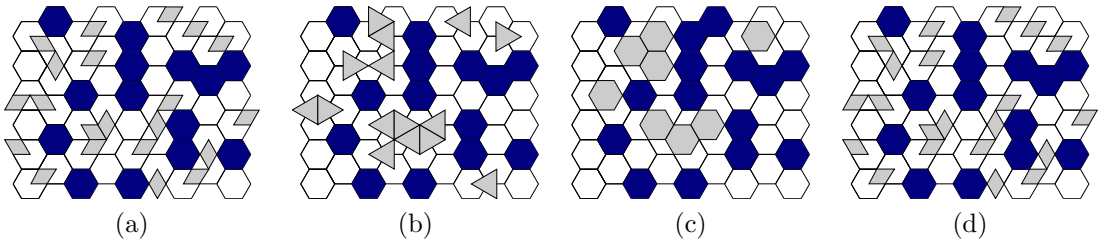


Figure 24: Models on the hexagonal lattice: (a) Ising Model, (b) Independent Sets, (c) Half-Independent Sets, (d) Half-Ising

well (which in turn increases the weight of the configuration). This would allow us to infer that all Interfering Binary Mixtures have the clustering property at high density. Define a component of A-tiles as a maximal connected subset, where x and y are adjacent if they share a vertex (and/or an edge). The number, $\mathcal{O}(\sigma)$, of open vertices in σ is equal to the total number of vertices minus the number of blocked vertices in σ. Thus for any component X, we want to bound the total number of B-faces blocked by X in terms of the perimeter of X.

For convenience, we will count the number $B(X)$ of B-faces blocked by a component X by counting the number $S(X)$ of s-shapes contained in blocked B-faces. Let α be the number of s-shapes making up a single A-tile, and β be the number of s-shapes making up a single B-tile. Then we obtain the number of blocked B-faces by noting that $B(X) = S(X)/\beta$. Notice that an s-shape in X is either contained within an A-tile of X or it is "hanging" off the boundary of X; let $H(X)$ be the set of *hanging* s-shapes of X, which we define as s-shapes contained in blocked B-faces but not contained in A-tiles of X. A hanging s-shape can either be incident to an edge of an A-tile in X or incident to a vertex of an A-tile (and not any edge).

For the Ising-type models, there is a direct correspondence between perimeter and the number of open B-vertices. In Figure 22d, we see an example of what we call a *half-Ising* model, where the B-tiles sit on half-edges of A-tiles. In this case, the perimeter of a component X is precisely twice the number of hanging s-shapes. However, in general, the shape of a component determines the exact relationship between its perimeter and the number of hanging s-shapes. Still, we can provide constant bounds on their ratio; let γ and δ be upper and lower bounds (respectively) on the ratio of the number of hanging s-shapes to the length of the perimeter of any component X, so that:

$$\delta\kappa(X) \leq H(X) \leq \gamma\kappa(X).$$

Note that γ and δ depend on the specific model. In the case of **Model 2**, $\gamma = 3$ suffices

Table 1: Parameters for the Interfering Binary Mixtures

	Fig. 22				Fig. 23				Fig. 24			
	a	b	c	d	a	b	c	d	a	b	c	d
α	4	4	2	8	3	3	6	2	6	6	3	12
β	2	4	4	2	2	6	2	6	2	3	3	2
γ	1	3	$\frac{3}{2}$	2	1	4	2	4	1	2	$\frac{1}{2}$	2
δ	1	1	$\frac{1}{2}$	2	1	1	2	1	1	1	$\frac{1}{2}$	2
ν	4	4	4	4	3	3	3	3	6	6	6	6

(in fact $\gamma = 2 + \frac{4}{\kappa(X)}$ works as well), and $\delta = 1$ since in the worst case (many holes) the number of hanging s-shapes is very close to the length of the perimeter. For each of the models depicted in Figures 22, 23, and 24, sufficient values for α, β, γ, and δ are given in Table 1 (note that in some cases we can find better values of γ and δ, but it is not our goal to get the best possible). In all cases, $\alpha|X| + \delta\kappa(X) \leq S(X) \leq \alpha|X| + \gamma\kappa(X)$, so we have the following bounds:

$$\frac{\alpha|X| + \delta\kappa(X)}{\beta} \;\leq\; B(X) \;\leq\; \frac{\alpha|X| + \gamma\kappa(X)}{\beta}. \tag{11}$$

Finally, let ν represent the perimeter of a tile in the A-lattice; this will be useful when we bound the maximum perimeter of a component in terms of its area. The values for ν are given in Table 1 as well.

5.5.3 Extending the analysis to other interfering binary mixtures

First we will show how to prove the analogs of Lemmas 5.3.1 and 5.3.2 for the triangular and hexagonal lattices. Next, we will show how to prove the analog of Theorem 5.2.1 for all interfering binary mixtures.

5.5.3.1 Bridge Systems and Finding a Sparse "Box"

Recall that in Sections 5.3 and 5.4, we consider the intersection of the A- and B-lattices with the finite region L, where L is an $n \times n$ grid. In order to simplify the comparison between different models, we will we assume that the number of A-faces

in L_A is n^2; for simplicity, the regions L we consider for the triangular and hexagonal lattices are the roughly rectangular regions pictured in Figure 25a and b. When we speak of an $l \times l$ "box" we mean either the $l \times l$ square grid, or the $\frac{l}{2} \times l$ region in Figure 25a, or the $\frac{l}{2} \times \frac{l}{2}$ region in Figure 25b, depending on the underlying A-lattice. As before, we assume we have bn^2 A-tiles.

Building Bridges. First we notice that the process for building bridge systems works the same way in each of our 12 models. As before, we will select horizontal lines through a given region and we consider the intersection of that line with the cells in that region. If the number of occupied cells along that line is at least c times the number of empty cells, then we add a set of bridges to the bridge system. For the triangular lattice, for every empty triangle intersecting the line H we add the edge parallel to H that intersects the bottom vertex or edge of that triangle, as in Figure 25a. For the hexagonal lattice, the bridges associated to the line H are the edges below the empty cells intersecting H which form a path as outlined in Figure 25b. Notice that these choices still allow us to encode the locations of all bridged components as desired. Moreover, the process outlined in the proof of Lemma 5.3.1 produces a c-bridge system (B, S) where the number of unbridged tiles is at most c times the number of unbridged cells, and $|B| \leq \kappa(S)(1-c)/(2c)$.

Finding a Sparse Box. We now show that after removing all but cn^2 tiles, there exists a roughly rectangular region of low density where we can place the tiles in our bank. Again the box we are looking for is of the shape described above.

Lemma 5.5.1. *For $(b-c)n^2 \leq a < bn^2$, there exists a constant $n_3 = n_3(b)$ such that for all $n \geq n_3$, if ρ is a configuration with at most cn^2 tiles then ρ contains a roughly rectangular region R' such that complementing R' requires a additional tiles and the change in total perimeter is at most $7\sqrt{a}$.*

Proof. Given a region R, let $d(R)$ denote the number of tiles needed to complement R;

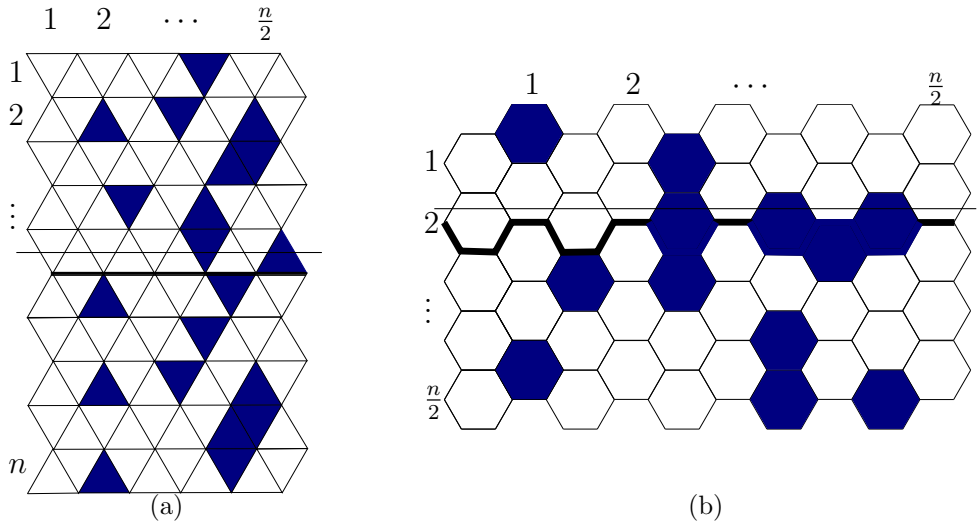


Figure 25: L-regions and $n \times n$ "boxes" for the triangular and hexagonal lattices ($n = 10$).

this is exactly the area of R (that is, the number of cells in R) minus twice the number of tiles in R. Let $l = \lceil \sqrt{8a/7} \rceil$. First we show that there exists an $l \times l$ box such that $d(R) \geq a$. Assume that such a box does not exist. Divide the A-lattice into $\left\lfloor \frac{n}{l} \right\rfloor^2$ disjoint $l \times l$ boxes, and consider any such box. Let t be the number of tiles in the box. The empty volume is at least $l^2 - t$. By assumption each box satisfies $l^2 - t < t + a$, and so $t > \frac{l^2 - a}{2}$. In particular, $8a/7 \leq l^2 < a + 2t \leq a + 2cn^2$, so we know $a < 14cn^2$. This implies that $l \leq \sqrt{8a/7} + 1 \leq 1 + 4\sqrt{c}n$. However, if T is the total number of tiles,

$$cn^2 \geq T > \left\lfloor \frac{n}{l} \right\rfloor^2 \frac{l^2 - a}{2} \geq \frac{n^2}{2} \left(1 - \frac{l}{n}\right)^2 \left(1 - \frac{a}{l^2}\right) > \frac{n^2 \left(1 - \frac{1}{n} - 4\sqrt{c}\right)^2}{16} \geq cn^2,$$

since $c \leq \frac{1}{65}$ and $n \geq n_3$, a contradiction. Therefore there exists an $l \times l$ box R such that $d(R) \geq a$. Remove cells from R one at a time, starting with the bottom row of R and moving across, until we obtain a region $R' \subseteq R$ with $d(R') = a$. This can be done because removing one cell at a time changes d by at most 1. This region R' is roughly rectangular and has perimeter at most $\nu\sqrt{8a/7} < 7\sqrt{a}$, since $\nu \leq 6$ for all models.

□

5.5.3.2 Proof of the General Theorem

Recall that our definition of the clustering property gives a precise bound on the maximum allowable perimeter for an active region that a configuration with the clustering property can have. However, the precise constant in that bound is not as essential as the fact that the perimeter is of order a square root of the order of the area of the region R. To extend the proof to all models on the grid, we must modify our definition of clustering slightly to accommodate the bounds we are using to relate the change in open vertices with the change in perimeter. Hence our definition of clustering becomes:

Definition 2. *We say that a configuration $\sigma \in \Omega$ (or $\Gamma(\sigma) \in \widehat{\Omega}$) has the* clustering property *if it contains a region R which satisfies the following properties:*

1. *R contains at least $(b-c)n^2$ A-tiles,*
2. *the perimeter of R is at most $14\frac{\gamma}{\delta}\sqrt{b}\ n$, and*
3. *the density of A-tiles in R is at least $1-c$ and in $\bar{R}$ is at most c.*

Theorem 5.5.2. *For any interfering binary mixture and for any $0 < b \leq 1/2$, there exist constants $\lambda^* = \lambda^*(b) > 1, \gamma_1 < 1$ and $n_1 = n_1(b)$ such that for all $n > n_1$, $\lambda \geq \lambda^*$ a random sample from Ω will have the clustering property with probability at least $(1 - {\gamma_1}^n)$.*

Proof. The proof proceeds as in the proof of Theorem 5.2.1. Let $\sigma \in \hat{\Omega} \backslash \Psi$. Construct a c-bridge system (B, S) for L_A as described in Lemma 5.3.1. Recall that for each bridged contour X with holes, we remove all outer tiles of $r(X)$ and complement all unbridged holes in X, using a subset of the tiles removed to fill in the holes. If $r(X)$ is a hole with islands, then we leave all of the unbridged islands alone. At this point, after complementing some number of regions, we have a bank of a extra tiles. Again,

by the definition of a c-bridge system, the density of tiles remaining is at most c, so $a \geq (b-c)n^2$.

Let $f_1(\sigma)$ be obtained from σ by removing the bridged components and complementing as described above. Let $\mathcal{F}_1$ be the image of f_1 on $\hat{\Omega} \setminus \Psi$. Let κ be the total perimeter of all contours bridged. Then as before, for any $\rho \in \mathcal{F}_1$, the number of preimages of ρ whose bridged contours have total perimeter κ is at most 5^{c_3} for $c_3 = (1 + \frac{1-c}{2c} + \frac{1}{14\sqrt{a}})\kappa$. Therefore $|f_1^{-1}(\rho)| \leq \sum_{\kappa \geq 14\sqrt{a}n} 5^{c_3}$.

Let $\rho \in \mathcal{F}_1$ with $bn^2 - a$ tiles. Lemma 5.5.1 shows how to find a region S' in ρ to complement using the a tiles from the bank to obtain τ in such a way that $\kappa(\tau) - \kappa(\rho) \leq 7\sqrt{a}$. Let $f_2(\rho) = \tau$ and $f = f_2 \circ f_1$. We can encode the boundary of S' with $n^2 3^{\kappa(S')} \leq n^2 3^{7\sqrt{a}}$ information. Hence for any $\tau \in \Psi$,

$$|f^{-1}(\tau)| \leq n^2 3^{7\sqrt{a}} \max_{\rho \in f_2^{-1}(\tau)} |f_1^{-1}(\rho)|.$$

Let $\sigma \in \hat{\Omega} \setminus \Psi$, and as above let κ be the total perimeter of components bridged in σ (recall $\kappa(\sigma)$ is the total perimeter of all contours in σ). Since $\sigma \notin \Psi$, we have $\kappa > 14\frac{\gamma}{\delta}\sqrt{a}$. Now, by equation 11, the change in the number of blocked vertices after applying the function f is at most

$$\left(\frac{\alpha a}{\beta} + \frac{\gamma\kappa(f(\sigma))}{\beta}\right) - \left(\frac{\alpha a}{\beta} + \frac{\delta\kappa(\sigma)}{\beta}\right) = \frac{-1}{\beta}\left(\delta\kappa(\sigma) - \gamma\kappa(f(\sigma))\right).$$

We have shown that

$$\delta\kappa(\sigma) - \gamma\kappa(f(\sigma)) \;>\; \delta\kappa - \gamma 5\sqrt{a} \;>\; \delta\kappa/2,$$

since $\kappa > 14\frac{\gamma}{\delta}\sqrt{a}$. Therefore the change $\mathcal{O}(\sigma) - \mathcal{O}(f(\sigma))$ in the number of open vertices is at least $\kappa\frac{\delta}{2\beta}$.

Let $\tau \in \Psi$ and define $f_\kappa^{-1}(\tau)$ to be the set of configurations with perimeter κ that map to τ. Then $|f_\kappa^{-1}(\tau)| \leq n^2 \left(3^{\frac{1}{2\sqrt{7}}} 5^{1+\frac{1-c}{2c}+\frac{1}{16\sqrt{b}}}\right)^\kappa$ and so

$$\pi(\tau)^{-1} \sum_{\sigma \in f^{-1}(\tau)} \pi(\sigma) \;\leq\; \sum_{\sigma \in f^{-1}(\tau)} \mu^{\mathcal{O}(\sigma)-\mathcal{O}(f(\sigma))} \;\leq\; \sum_{\kappa=14\frac{\gamma}{\delta}\sqrt{a}}^{2n^2} \mu^{\kappa\delta/(2\beta)} |f_\kappa^{-1}(\tau)| \;\leq \gamma_1^n,$$

for some $\gamma_1 < 1$, if $\mu \leq \mu^* < \left(3^{\frac{1}{2\sqrt{7}}} 5^{1+\frac{1-c}{2c}+\frac{1}{16\sqrt{b}}}\right)^{-2\beta/\delta}$. Thus the theorem holds if $\lambda \geq \lambda^* = \mu^{*-2} - 1$.

□

5.6 Conclusion

The clustering of the Ising model has been studied extensively (see, e.g. [21], or the references therein), down to the precise shape of the boundary of the largest "cluster", known as a Wulff shape. While it was believed that other models behaved similarly to the Ising model, the heavy machinery developed for the Ising model had previously been too specialized to apply in more general settings. In this chapter, we provide much simpler arguments to explain the clustering behavior of the Ising model at high density that also extend to other interfering binary mixtures.

In addition, we show that each of these models of colloids fits into a framework of independent sets on some graph. In particular, we show that the Ising model on a graph G can be viewed as independent sets on a different graph G'. One area of future research is to identify other graphs (besides the 12 given here) for which we can characterize the clustering of independent sets on those graphs.

REFERENCES

[1] Aldous, D., "Random walk on finite groups and rapidly mixing Markov chains," *In Seminaire de Probabilites XVII*, pp. 243–297, 1983.

[2] Aldous, D. and Diaconis, P., "Shuffling cards and stopping times," *Amer. Math. Monthly*, vol. 93, pp. 333–348, 1986.

[3] Alon, N., "Eigenvalues and expanders," *Combinatorica*, vol. 6, pp. 83–96, 1986.

[4] Alon, N. and Milman, V., "λ_1, isoperimetric inequalities for graphs and superconcentrators," *Journal of Combinatorial Theory Series B*, vol. 38, pp. 73–88, 1985.

[5] Benjamini, I., Berger, N., Hoffman, C., and Mossel, E., "Mixing times of the biased card shuffling and the asymmetric exclusion process," *Trans. Amer. Math. Soc*, vol. 357, pp. 3013–3029, 2005.

[6] Bhakta, P., Miracle, S., Streib, A., and Randall, D., "Mixing times of self-organizing lists and biased permutations.," *Submitted.*

[7] Bhatnagar, N. and Randall, D., "Torpid mixing of simulated tempering on the potts model," in *Proceedings of the 15th ACM/SIAM Symposium on Discrete Algorithms*, SODA '04, pp. 478–487, 2004.

[8] Birdi, K., *Handbook of Surface and Colloid Chemistry.* CRC Press, 2008.

[9] Borgs, C., Chayes, J., Frieze, A., Kim, J., Tetali, P., Vigoda, E., and Vu, V., "Torpid mixing of some MCMC algorithms in statistical physics," FOCS 1999, 1999.

[10] Borgs, C., Chayes, J., Frieze, A., Kim, J., Tetali, P., Vigoda, E., and Vu, V., "Torpid mixing of some mcmc algorithms in statistical physics," in *40th IEEE Symp. on Foundations of Computer Science*, FOCS 1999, pp. 218–229, 1999.

[11] Bubley, R. and Dyer, M., "Faster random generation of linear extensions," in *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, SODA '98, 1998.

[12] Bubley, R. and Dyer, M., "Faster random generation of linear extensions.," *Discrete Math.*, vol. 201, pp. 81–88, 1999.

[13] Buhot, A. and Krauth, W., "Phase separation in two-dimensional additive mixture," *Phys. Rev.*, vol. E 59, pp. 2939–2941, 1999.

[14] Cesi, F., Guadagni, G., Martinelli, F., and Schonmann, R., "On the two-dimensional dynamical Ising model in the phase coexistence region near the critical point," *Journal of Statistical Physics*, vol. 85, pp. 55 –102, 1996.

[15] Chayes, J., Chayes, L., and Schonmann, R., "Exponential decay of connectivities in the two dimensional Ising model," *Journal of Statistical Physics*, vol. 49, pp. 433 –445, 1987.

[16] Cook, M., Gacs, P., and Winfree, E., "Self-stabilizing synchronization in 3 dimensions," *Preprint*, 2008.

[17] Diaconis, P., *Group representations in probability and statistics*, vol. 11 of *Lecture Notes – Monograph Series*. Institute of Mathematical Statistics, 1988.

[18] Diaconis, P. and Saloff-Coste, L., "Comparison theorems for reversible markov chains," *The Annals of Applied Probability*, vol. 3, pp. 696–730, 1993.

[19] Diaconis, P. and Shahshahani, M., "Generating a random permutation with random transpositions," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 57, pp. 159–179, 1981.

[20] Dobrushin, R., "The problem of uniqueness of a Gibbs random fields and the problem of phase transition," *Func. anal. and appl.*, vol. 2, pp. 302–312, 1968.

[21] Dobrushin, R., Kotecký, R., and Shlosman, S., *The Wulff Construction: A Global Shape from Local Interactions*. American Mathematical Society, 1992.

[22] Dobrushin, R. and Shlosman, S., "Constructive criterion for the uniqueness of Gibbs fields," in *Statistical Physics and Dynamical Systems*, pp. 347–370, Birhauser, 1985.

[23] Dress, C. and Krauth, W., "Cluster algorithm for hard spheres and related systems," *J. Phys. A: Math. Gen.*, vol. 28, 1995.

[24] Dyer, M. and Greenhill, C., "A more rapidly mixing markov chain for graph colorings," *Random Structures & Algorithms*, vol. 13, pp. 285–317, 1998.

[25] Dyer, M., Frieze, A., and Kannan, R., "A random polynomial time algorithm for approximating the volume of convex bodies," *Journal of the ACM*, vol. 38, pp. 1–17, 1991.

[26] Feller, W., *An introduction to probability theory and its applications*. Wiley, 1968.

[27] Fill, J., "Background on the gap problem," *Unpublished manuscript*, 2003.

[28] Fill, J., "An interesting spectral gap problem," *Unpublished manuscript*, 2003.

[29] Frenkel, D. and Louis, A., "Phase separation in binary hard-core mixtures: An exact result.," *Physical Review Letters*, vol. 68, pp. 3363–3365, 1992.

[30] Fu, T.-J. and Seeman, N., "DNA double-crossover molecules," *Biochemistry*, vol. 32, pp. 3211 – 3220, 1993.

[31] Geman, S. and Geman, D., "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. PAMI*, vol. 6, pp. 721–741, 1984.

[32] Gore, V. and Jerrum, M., "The Swendsen-Wang algorithm does not always mix rapidly," *J. Statistical Phys.*, vol. 97, pp. 67–86, 1999.

[33] Greenberg, S., Pascoe, A., and Randall, D., "Sampling biased lattice configurations using exponential metrics," in *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '09, 2009.

[34] Hastings, W., "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[35] Heilmann, O. and Lieb, E., "Theory of monomer-dimer systems," *Communications in Mathematical Physics*, vol. 25, pp. 190–232, 1972.

[36] Hester, J. H. and Hirschberg, D. S., "Self-organizing linear search," *Computing Surveys*, vol. 17, pp. 295–311, 1985.

[37] Hiemenz, P. and Rajagopalan, R., *Principles of Colloid and Surface Chemistry*. CRC Press, 1997.

[38] Jerrum, M., *Counting, sampling and integrating: Algorithms and complexity*. Lectures in Mathematics ETH Zurich, 2003.

[39] Jerrum, M. and Sinclair, A., "Conductance and the rapid mixing property for Markov chains: the approximation of the permanent resolved," Proc. 20nd Annual ACM Symposium on Theory of Computing, pp. 235–243, 1988.

[40] Jerrum, M. and Sinclair, A., "Approximate counting, uniform generation and rapidly mixing markov chains," *Information and Computation*, vol. 82, pp. 93–133, 1989.

[41] Jerrum, M. and Sinclair, A., "Polynomial-time approximation algorithms for the Ising model," *SIAM Journal on Computing*, vol. 22, pp. 1087–1116, 1993.

[42] Jerrum, M. and Sinclair, A., "The Markov chain Monte Carlo method: an approach to approximate counting and integration," in *Approximation Algorithms for NP-hard Problems* (Hochbaum, D., ed.), PWS, 1996.

[43] Jerrum, M., Valiant, L., and Vazirani, V., "Random generation of combinatorial structures from a uniform distribution," *Theoretical Computer Science*, vol. 243, pp. 169–188, 1986.

[44] Jerrum, M., Sinclair, A., and Vigoda, E., "A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries," *Journal of the ACM*, vol. 51, pp. 671–697, 2004.

[45] Kannan, R., Lovász, L., and Simonovits, R., "Random walks and an $o^*(n^5)$ volume algorithm for convex bodies," *Random Structures and Algorithms*, vol. 11, pp. 1–50, 1987.

[46] Lawler, G. and Sokal, A., "Bounds on the l^2 spectrum for markov chains and markov processes: a generalization of cheeger's inequality," *Transactions of the American Mathematical Society*, vol. 309, pp. 557–580, 1988.

[47] Levin, D. A., Peres, Y., and Wilmer, E. L., *Markov chains and mixing times*. American Mathematical Society, 2006.

[48] Liu, J., *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics, Springer, 2004.

[49] Luby, M., Randall, D., and Sinclair, A., "Markov chains for planar lattice structures," *SIAM Journal on Computing*, vol. 31, pp. 167–192, 2001.

[50] Majumder, U., Sahu, S., and Reif, J., "Stochastic analysis of reversible self-assembly.," *Journal of Computational and Theoretical Nanoscience*, To appear, 2008.

[51] Markov, A., "Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga," *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete, 2-ya seriya*, vol. tom 15, pp. 135 – 156, 1906.

[52] Martinelli, F., "Lectures on Glauber dynamics for discrete spin models," *Lectures Notes in Math*, vol. 1717, pp. 93–191, 1999.

[53] Martinelli, F. and Olivieri, E., "Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case," *Comm. Math. Phys.*, vol. 161, no. 3, pp. 447–486, 1994.

[54] Martinelli, F. and Olivieri, E., "Approach to equilibrium of Glauber dynamics in the one phase region. II. The general case," *Comm. Math. Phys.*, vol. 161, no. 3, pp. 487–514, 1994.

[55] McShine, L. and Tetali, P., "On the mixing time of the triangulation walk and other catalan structures," in *DIMACS-AMS volume on Randomization Methods in Algorithm Design* (et al., P., ed.), vol. 43, pp. 147–160, 1998.

[56] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and E., T., "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, pp. 1087–1092, 1953.

[57] Miracle, S., Randall, D., and Streib, A. P., "Cluster algorithms for discrete models of colloids with bars," in *8th Workshop on Analytic Algorithmics and Combinatorics*, ANALCO, pp. 135–150, 2011.

[58] Miracle, S., Randall, D., and Streib, A. P., "Clustering in interfering binary mixtures," in *14th International Workshop on Randomization and Computation*, RANDOM, pp. 652–663, 2011.

[59] Papadimitriou, C., *Computational Complexity*. Addison Wesley, 1993.

[60] Pascoe, A. and Randall, D., "Self-assembly and convergence rates of heterogeneous reversible growth processes," *6th Foundations of Nanoscience*, vol. 38, pp. 48–59, 2009.

[61] Randall, D., "Slow mixing of glauber dynamics via topological obstructions," in *17th ACM-SIAM Symp. on Discrete Alg*, SODA, pp. 870–879, 2006.

[62] Randall, D. and Tetali, P., "Analyzing glauber dynamics by comparison of markov chains," *Journal of Mathematical Physics*, vol. 41, pp. 1598–1615, 2000.

[63] Randall, D. and Tetali, P., "Analyzing glauber dynamics by comparison of markov chains," in *Proceedings of the Third Latin American Symposium on Theoretical Informatics*, LATIN '98, pp. 292–304, 1998.

[64] Seeman, N., "DNA in a material world," *Nature*, vol. 421, pp. 427–431, 2003.

[65] Thomas, L., "Bound on the mass gap for finite volume stochastic Ising models at low temperature," *Communications in Mathematical Physics*, vol. 126, pp. 1–11, 1989.

[66] Valiant, L., "The complexity of computing the permanent," *Theoretical Computer Science*, vol. 8, pp. 189–201, 1979.

[67] Wilson, D., "Mixing times of lozenge tiling and card shuffling markov chains," *The Annals of Applied Probability*, vol. 1, pp. 274–325, 2004.

[68] Winfree, E., "Simulations of computing by self-assembly," in *4th DIMACS Meeting on DNA Based Computers*, 1998.

[69] Winfree, E., Yang, X., and Seeman, N., "Universal computation via self-assembly of DNA: Some theory and experiments.," *DNA Based Computers II*, vol. 44, pp. 191–213, 1996.